Dimensionality reduction and pathway network analysis of transcriptome data: Application to T-cell characterization

Soutenance de Thèse de Christophe Bécavin

6 Décembre 2010

Rapporteurs: Alain Arneodo, Laboratoire Joliot-Curie, ENS Lyon
Lars Rogge, Institut Pasteur, Paris

Examinateurs: Jean-Marc Victor, LPTMC, Paris 6
Sylviane Pied, Institut Pasteur, Lille
Andrei Zinovyev, Institut Curie, Paris

Directrice de thèse: Annick Lesne, LPTMC, Paris 6
Co-Directeur de thèse: Arndt Benecke, IHES, Paris

Transcriptome

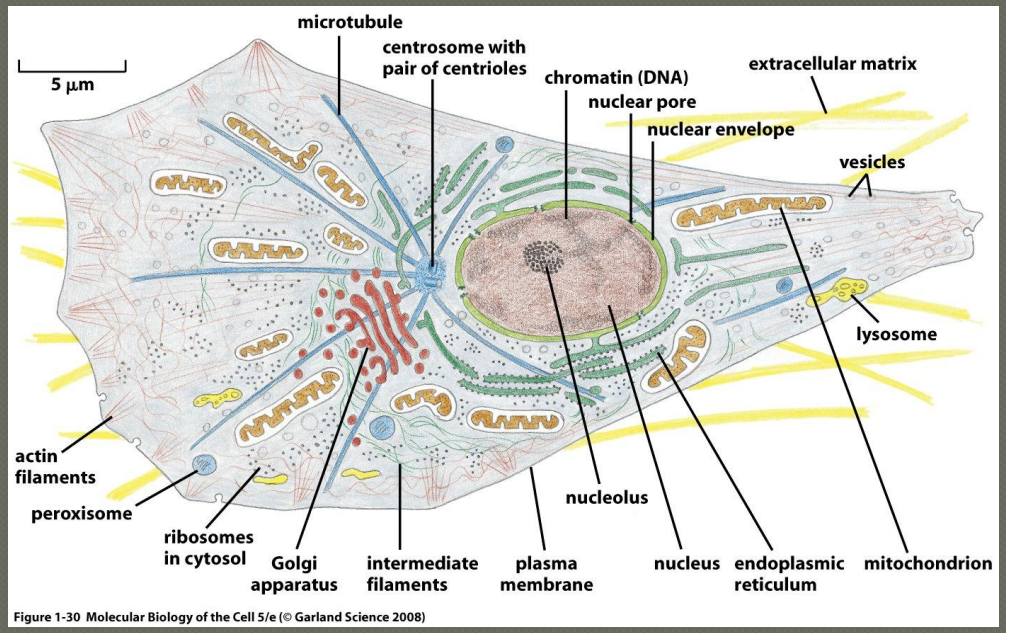


Figure 1-30 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Albert et al., Molecular Biology of the Cell, 4th Ed

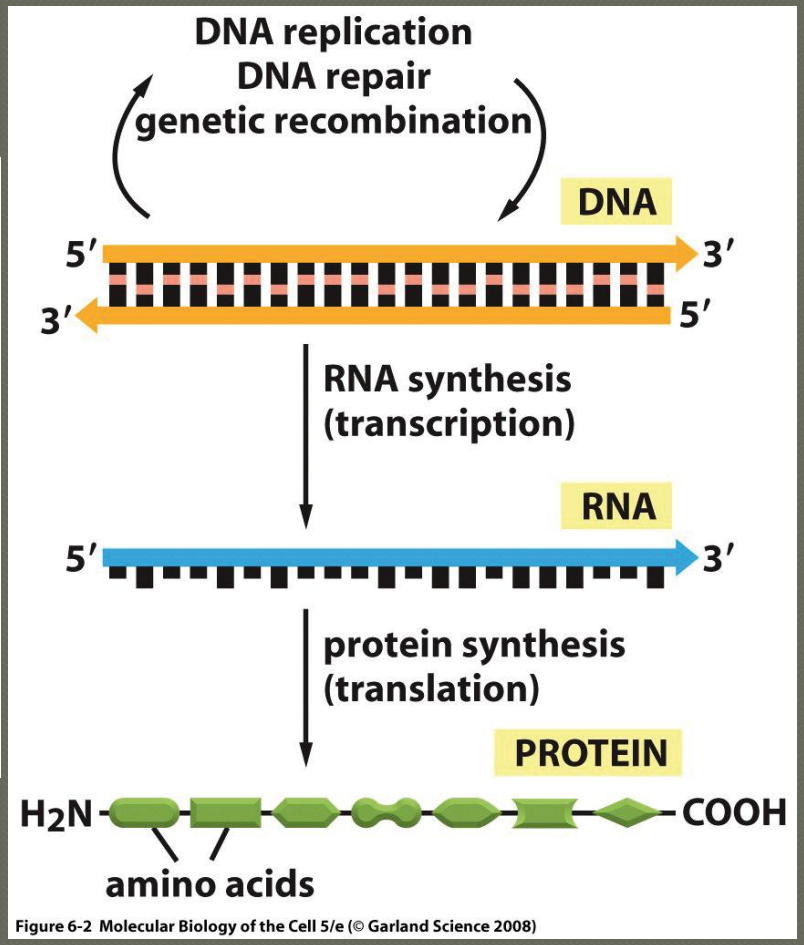
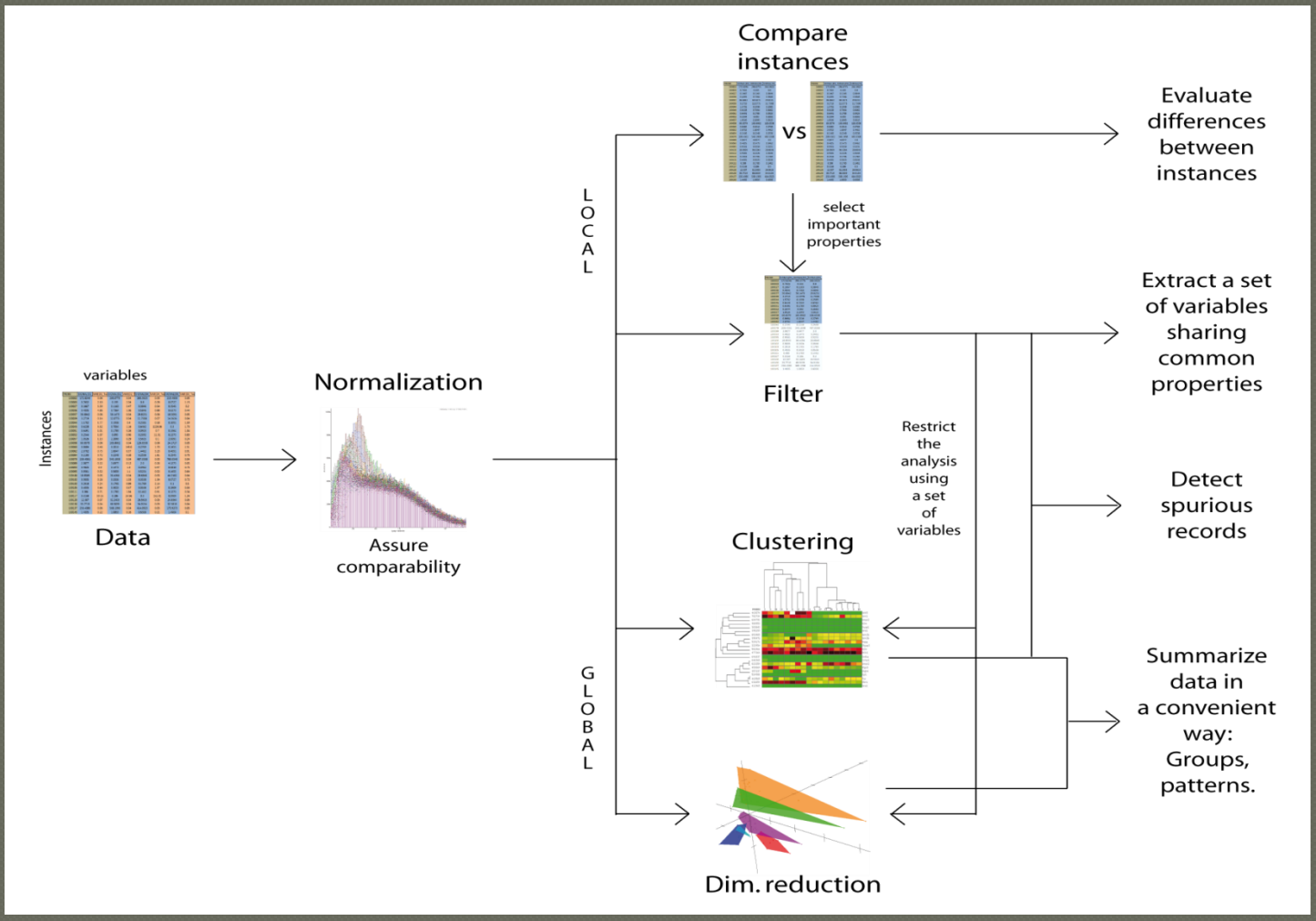


Figure 6-2 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Typical work-flow in data analysis



Index

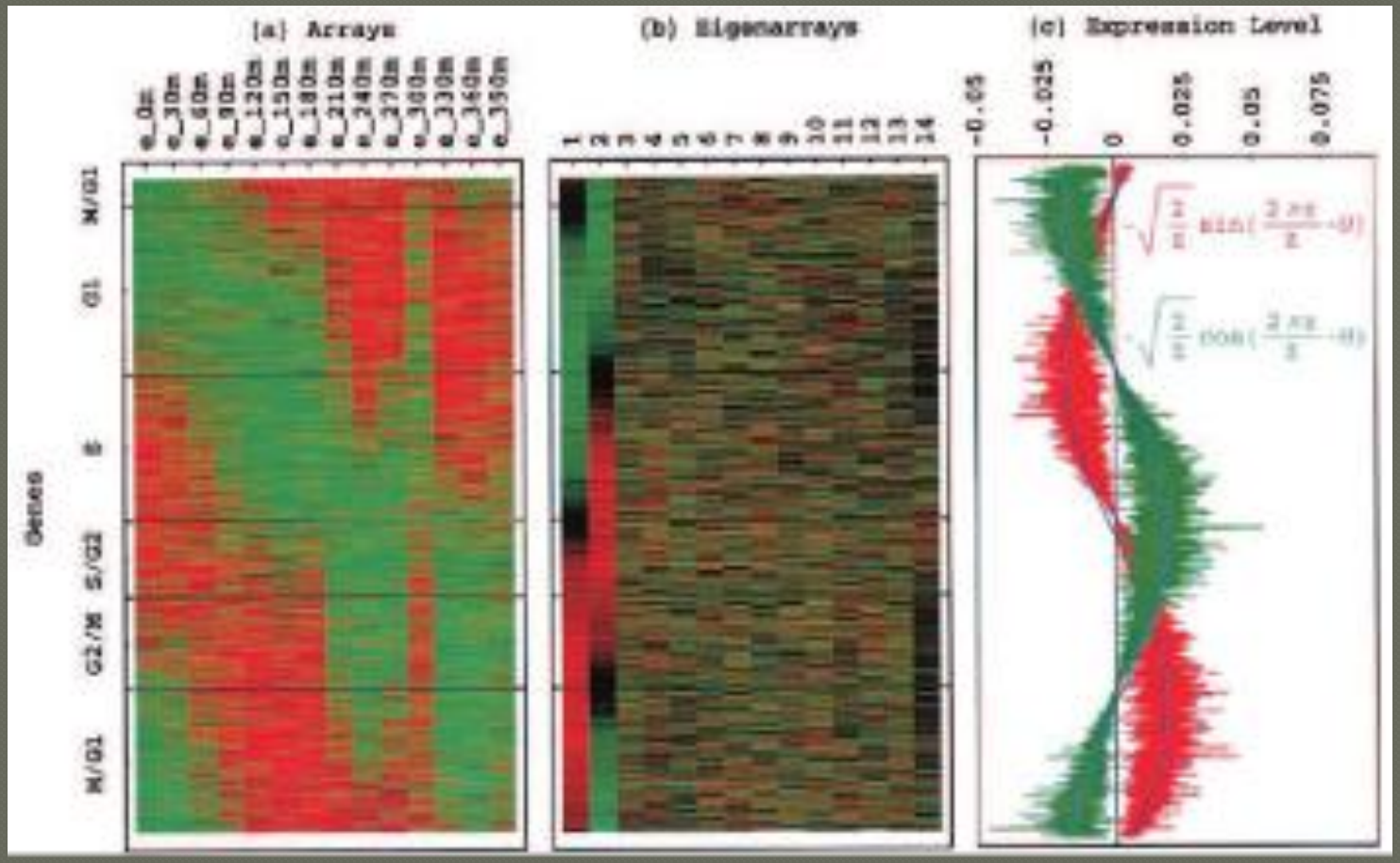
- Review of dimensionality reduction techniques
- SVD-MDS
- Pathway network analysis software
- 1st Application: Cerebral Malaria characterization
- 2nd Application: Treg characterization

Dimensionality reduction

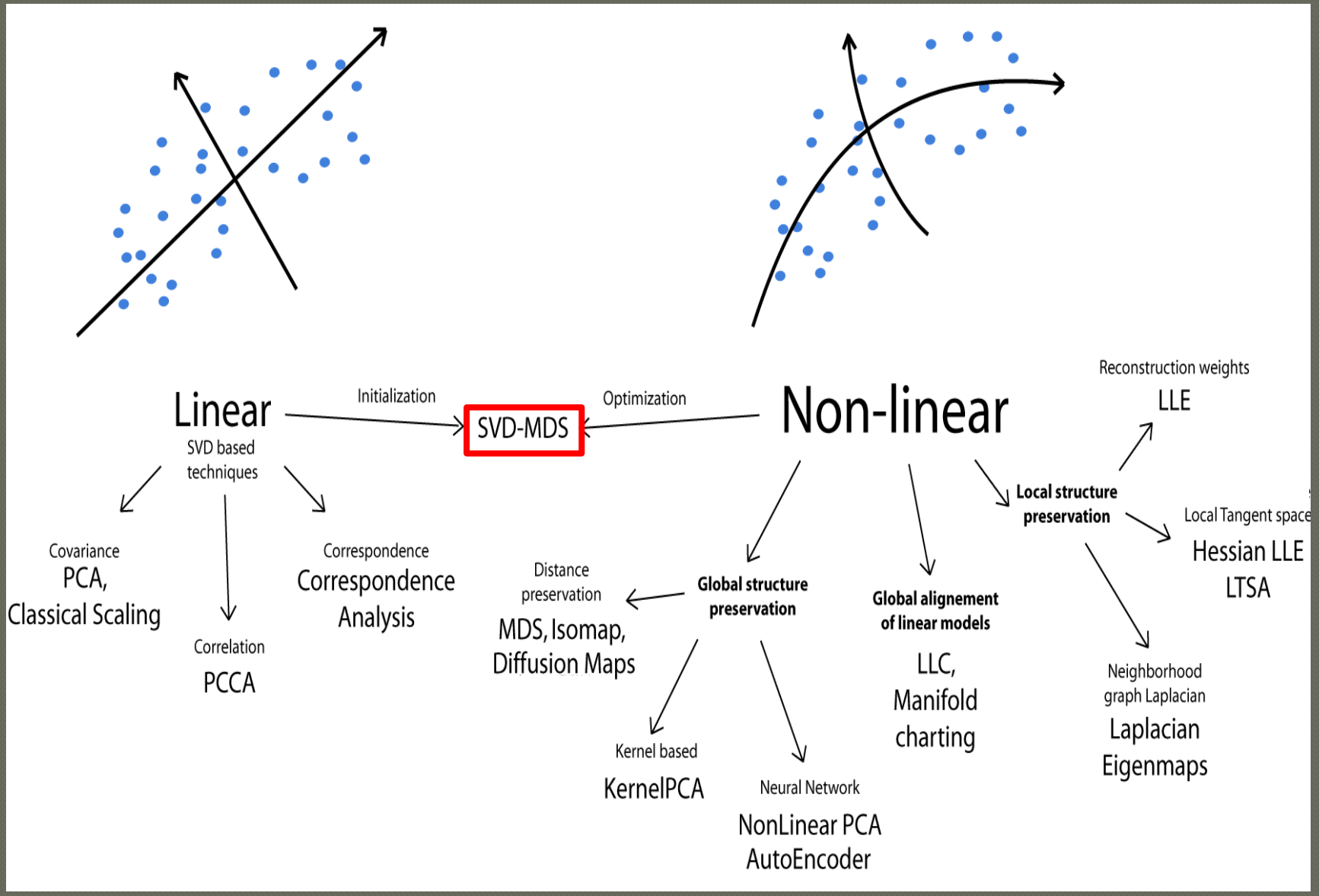
PROBE	SIGNAL[D1_	VAR[D1_Tcc	SIGNAL[D2_	VAR[D2_T	SIGNAL[D4_	VAR[D4_Tcc	SIGNAL[D5_	VAR[D5_Tcc
100002	172.8246	0.06	295.0775	0.04	263.3923	0.05	215.4909	0.05
100003	0.7632	2.19	0.155	2.04	0.8	0.29	0.0727	1.15
100027	0.1667	5.39	0.1163	3.47	0.3846	0.54	0.3545	0.2
100036	0.4035	4.86	0.7364	1.36	0.5846	0.89	0.6273	0.44
100037	36.6842	0.06	58.1473	0.04	29.8231	0.06	18.5091	0.05
100039	3.2719	0.14	12.0775	0.04	11.7308	0.07	14.3636	0.04
100044	1.5702	1.77	0.1938	0.4	0.2385	0.58	0.1091	1.59
100045	0.6228	5.52	0.7054	1.16	0.6462	1228.36	0.3	1.73
100051	0.6491	1.01	0.1783	3.26	0.0923	0.7	0.1364	1.04
100052	0.2544	1.07	0.093	0.96	0.2692	11.31	0.2273	0.83
100057	1.0526	1.13	2.2093	0.29	3.5615	0.1	2.0091	0.24
100058	60.6579	0.06	165.8062	0.04	126.6538	0.06	24.1727	0.05
100060	0.8684	0.46	0.2016	149.6	0.2769	1.73	0.1455	2.51
100062	2.0702	0.75	1.6047	6.37	1.4462	3.23	0.4091	0.91
100064	0.2193	3.72	0.2248	0.26	0.2538	1.81	0.2545	0.75
100079	209.4561	0.04	540.1938	0.04	487.0308	0.05	769.0546	0.04
100089	2.0877	0.22	3.6977	0.12	2.0	0.26	4.1273	0.05
100093	0.4825	0.9	0.1473	1.3	0.3462	0.47	0.0818	0.75
100095	0.9561	0.52	0.5659	1.1	0.5231	0.52	0.1455	0.64
100100	18.8509	0.05	36.4264	0.04	28.6846	0.05	44.3182	0.04
100102	0.9035	0.26	0.2326	1.03	0.3538	1.34	0.0727	0.72
100103	0.2018	4.24	0.1705	0.89	0.1769	2.14	0.1	0.5
100105	0.4035	3.64	0.3023	0.57	0.0538	1.37	0.1909	0.55
100111	0.386	0.71	0.1783	1.84	0.1462	0.91	0.1273	0.56
100117	0.3158	19.15	0.186	24.56	0.3	152.31	0.0909	1.24
100120	12.307	0.07	32.2403	0.04	26.5615	0.05	24.8364	0.05
100130	35.7719	0.04	68.5039	0.04	54.9154	0.05	32.5818	0.04
100137	256.4386	0.06	568.1396	0.04	414.0923	0.05	173.9273	0.05
100145	1.4035	0.22	1.0853	0.16	0.6308	0.21	1.4909	0.1

PROBE	axe 1	axe 2	axe 3
100002	-0,005761155	-0,0196512	0,0119491
100003	3,33E-06	-7,64E-06	-3,41E-05
100027	1,56E-05	1,68E-05	-1,35E-05
100036	-2,15E-05	5,47E-05	7,82E-05
100037	-6,07E-04	-0,0032565	-3,98E-04
100039	-1,81E-04	-8,52E-04	7,59E-04
100044	-7,69E-06	1,92E-05	-4,63E-05
100045	-1,48E-06	2,01E-05	-1,08E-05
100051	2,72E-05	1,03E-05	-1,29E-05
100052	2,45E-05	2,98E-05	-4,44E-07
100057	-3,12E-05	9,81E-06	8,40E-05
100058	-0,005544852	-0,0014884	3,56E-05
100060	-4,28E-06	2,66E-05	-2,50E-05
100062	-1,87E-05	1,36E-04	8,07E-05
100064	1,16E-05	1,48E-05	-1,36E-06
100079	-0,008017877	-0,052928	0,0132812
100089	-3,44E-05	-7,79E-05	-6,46E-05
100093	1,21E-05	1,12E-05	-3,78E-05
100095	-2,46E-06	2,76E-05	-4,02E-05
100100	-4,15E-04	-1,19E-04	2,03E-04
100102	4,28E-06	1,75E-05	-3,19E-05
100103	8,46E-06	1,68E-05	1,43E-05
100105	3,09E-05	5,10E-05	5,02E-05
100111	8,07E-06	1,77E-05	-6,58E-06
100117	9,05E-06	3,10E-05	-5,17E-05
100120	-8,48E-04	-0,0027472	5,07E-04
100130	-0,001304086	-0,0035659	3,79E-04
100137	-0,018473712	3,53E-04	0,015201
100145	1,55E-05	-7,17E-06	-3,92E-05

Applications of dimensionality reduction techniques



Alter et al., PNAS, 2000

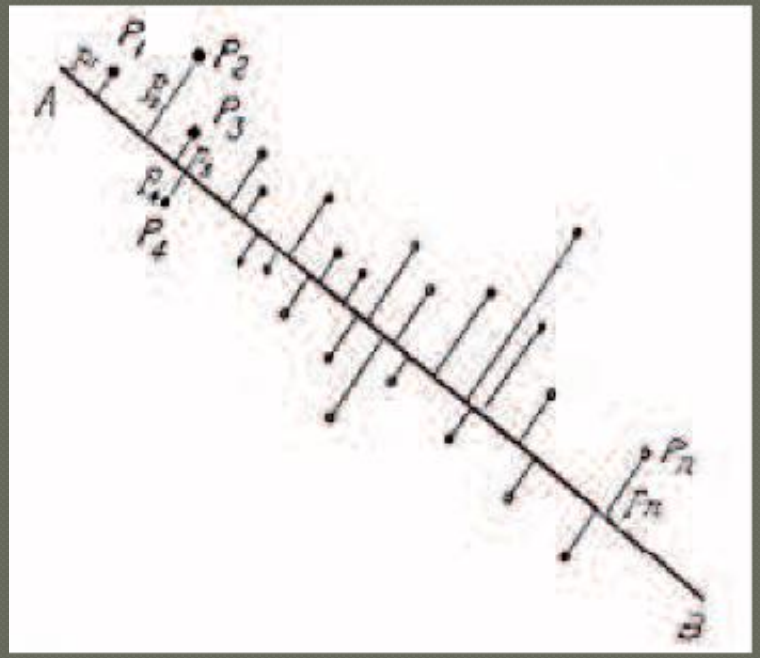


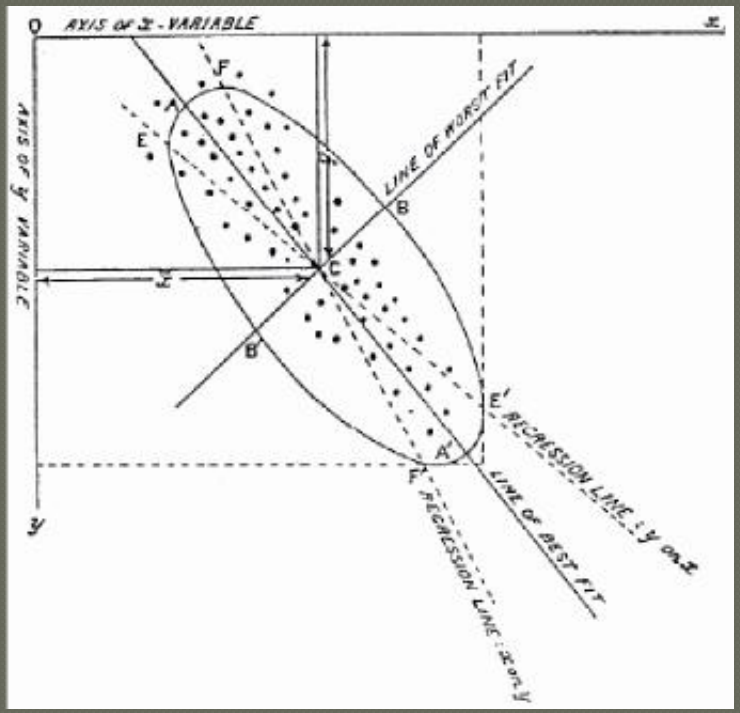
Bécavin et al., Expert Review of Molecular Diagnostics, 2011

Principal Component Analysis

In many physical, statistical, and biological investigations it is desirable to represent a system of points in a plan, three, or higher dimensioned space by the "best-fitting" straight line or plane

K. Pearson, Philosophical Magazine, 1901





Different measures for different linear dimensionality reduction techniques

Covariance (PCA):

$$Cov(x_i, x_j) = E\left[(x_i - E[x_i])(x_j - E[x_j])\right]$$

$$Cov(x_i, x_j) = 1/n \sum_{k=1}^{n} x_i^k x_j^k$$

$$CovMatrix = 1/n \, X^t X$$

Correlation (PCCA):

$$Cov(x_i, x_j) = \frac{1/n \sum_{k=1}^{n} x_i^k x_j^k}{\sigma(x_i)\sigma(x_j)}$$

$$\tilde{X} = \left(\frac{x_{i,j}}{\sigma(j)} \right)$$

Euclidean distance (Classical scaling):

$$d(x^i, x^j)^2 = \sum_k \left(x_k^i - x_k^j \right)^2$$

$$XX^t = -\frac{1}{2} H \begin{pmatrix} d_{11}^{\,2} & d_{12}^{\,2} & \vdots \\ d_{21}^{\,2} & \ddots & \vdots \\ \cdots & \cdots & d_{nn}^{\,2} \end{pmatrix} H$$

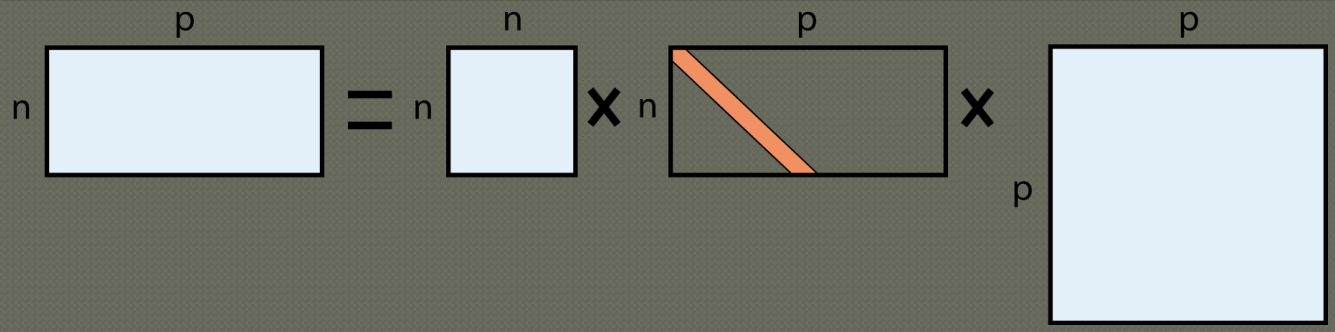
$$with \quad H = I - 1/n \, 11^t, and \quad 1 = (1,1,\cdots,1)^t$$

Chi-square distance (correspondence analysis):

$$\chi^2(x^i, x^j)^2 = \sum_k \left(\frac{x_k^i}{\sqrt{W \sum_l x_k^l} \cdot \sum_l x_l^i} - \frac{x_k^j}{\sqrt{W \sum_l x_k^l} \cdot \sum_l x_l^j} \right)^2$$

$$with \quad W = \sum_k \sum_l x_l^k$$

Singular Value Decomposition



$$X = USV^t$$

with $s_1 > s_2 > \cdots > s_{n-1} > 0$

Inner product matrix

$$XX^t = U\lambda U^t$$

Outer product matrix

$$X^t X = V\lambda V^t$$

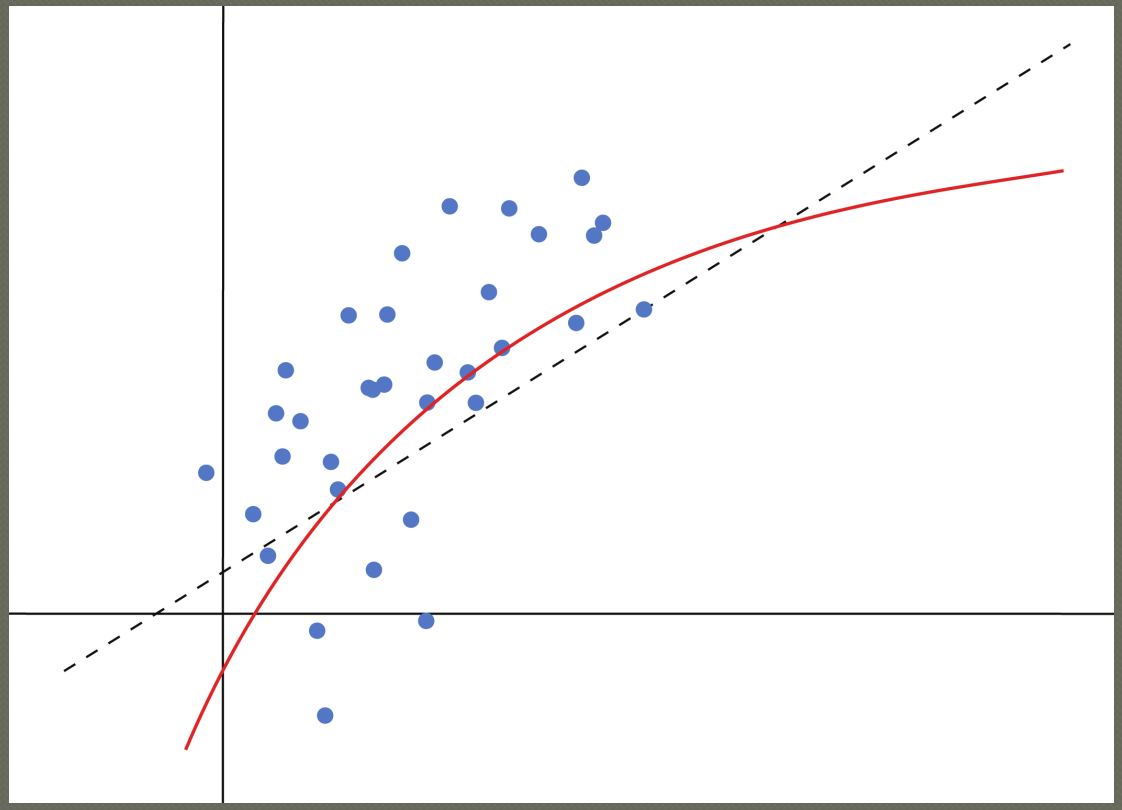
It is easy to demonstrate that inner and outer product matrices have the same eigenvalues, which are linked to singular values

$$s_i^{\,2} = \lambda_i$$

From line to curve: Non-linear PCA

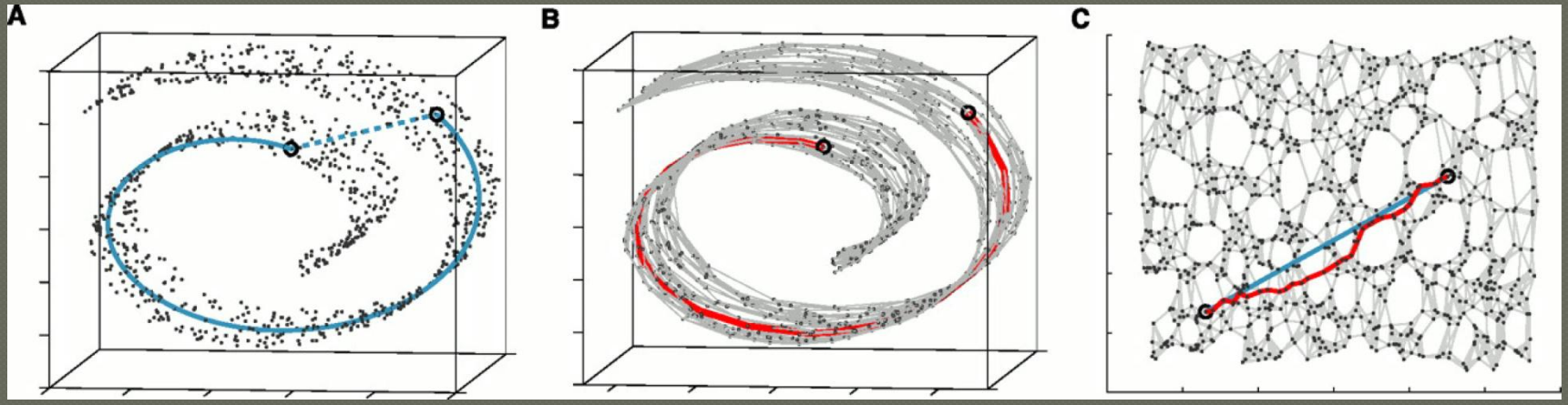
PCA
search for the "best-fitting" line

Non-linear PCA
search for the principal curve



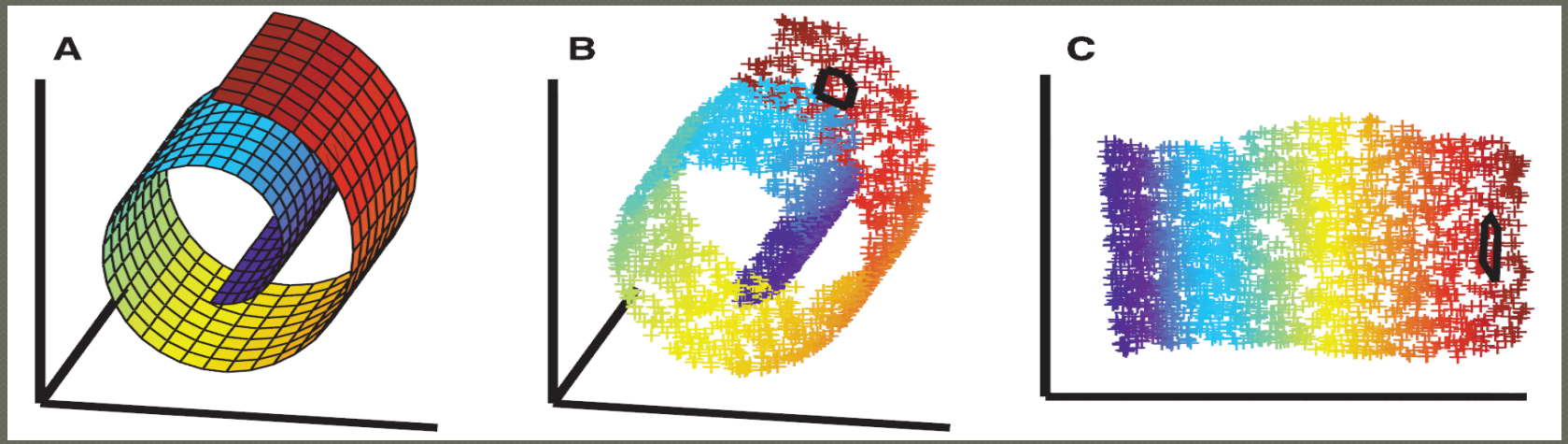
Isomap

Tenenbaum et al. , Science, 2000

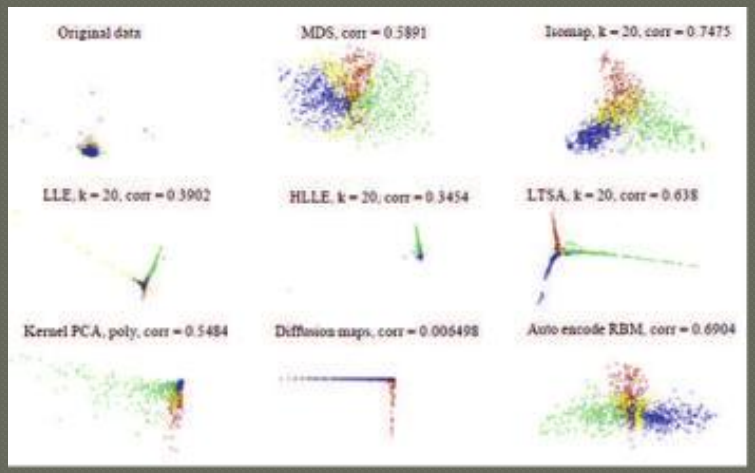


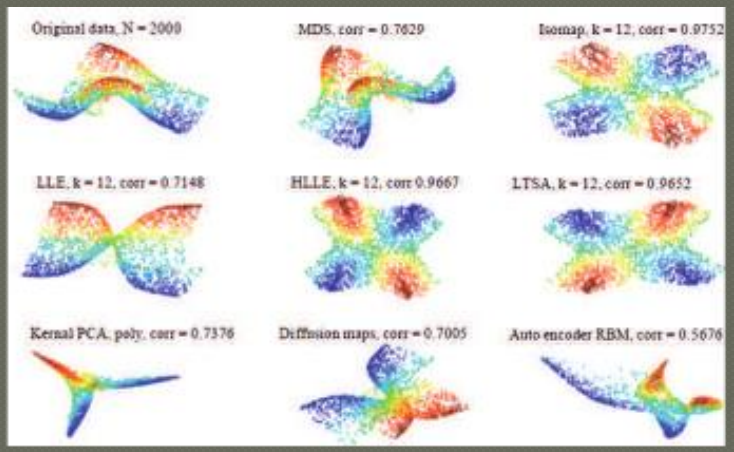
Locally Linear Embedding

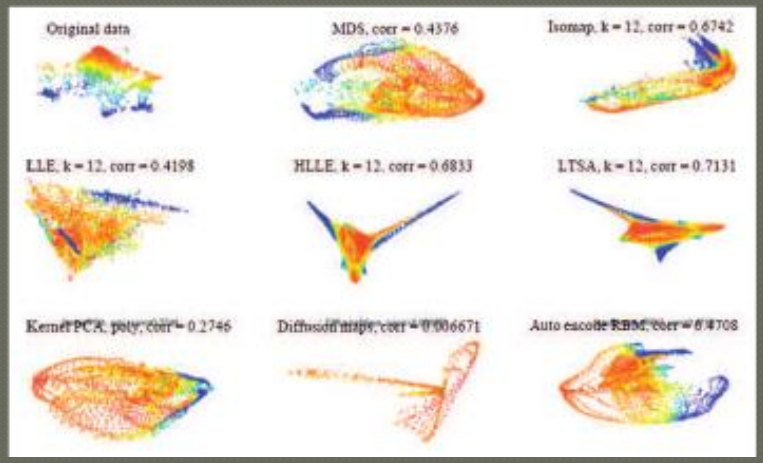
Roweis et al. , Science, 2000

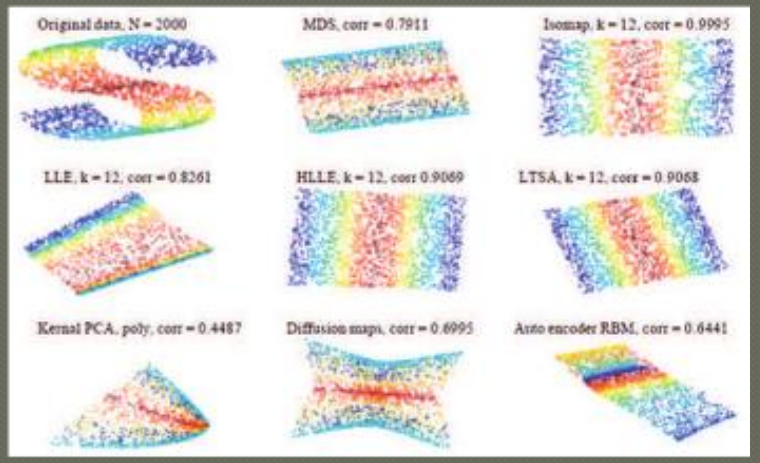


A great diversity of results given by each dimensionality reduction technique









Tsai F., Journal of Artificial Intelligence, 2010

Multidimensional Scaling

using Molecular Dynamics approach

-Each pair of points connected by a spring:

$$F_{spring}\left(x^i, x^j\right) = -K.\left(\delta\left(x^i, x^j\right) - d\left(x^i, x^j\right)\right)^2$$

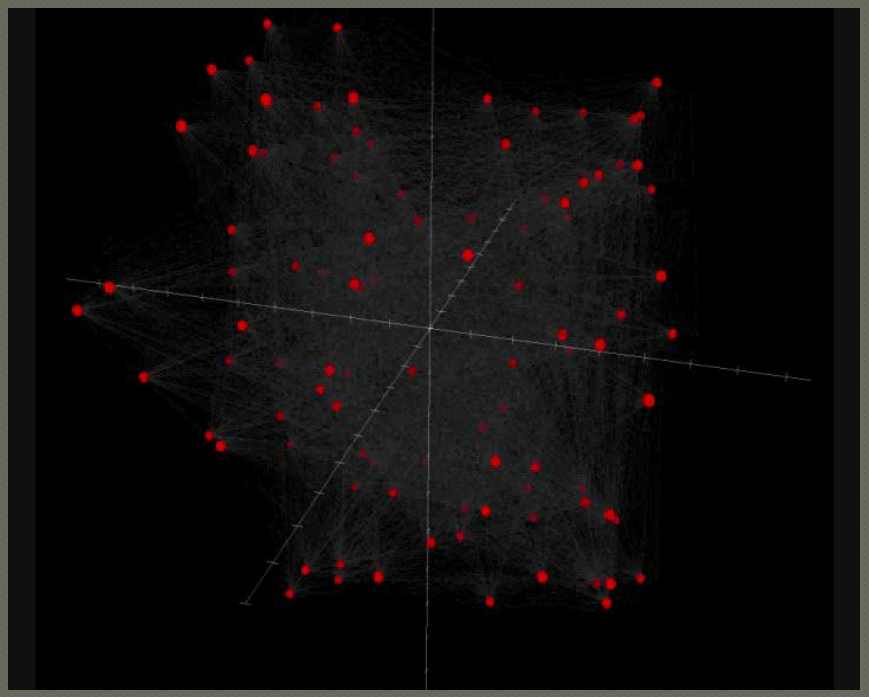
- Calculation of force for all points (non-conservative system):

$$F(x^i) = \sum_{k \neq i} F_{spring}\left(x^i, x^k\right) - \gamma \cdot mass \cdot \dot{x}^i$$

-Find new position of all points using Newton equation, and Verlet integration:

$$mass \cdot \ddot{x}^i = F(x^i)$$

$$x^i(t + \Delta t) = 2x^i(t) - x^i(t - \Delta t) + \ddot{x}^i \Delta t^2$$

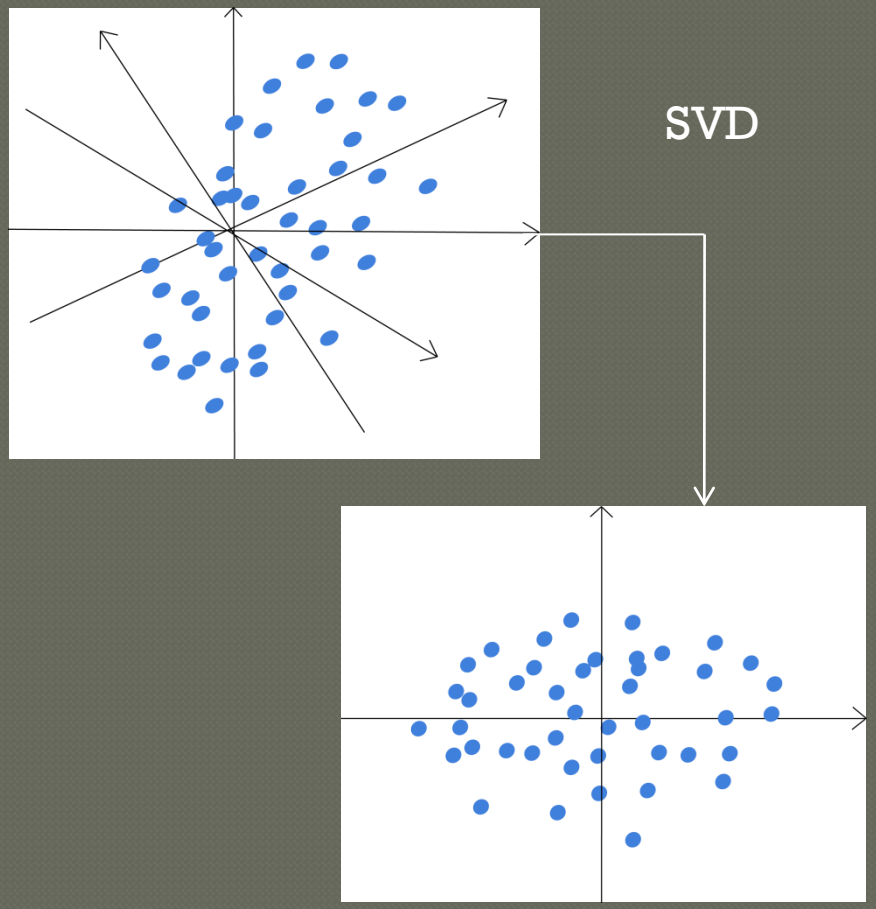


$$Stress = \sqrt{\frac{\sum_i \sum_j \left(\delta(i,j) - d(i,j)\right)^2}{\sum_i \sum_j \left(d(i,j)\right)^2}}$$

Index

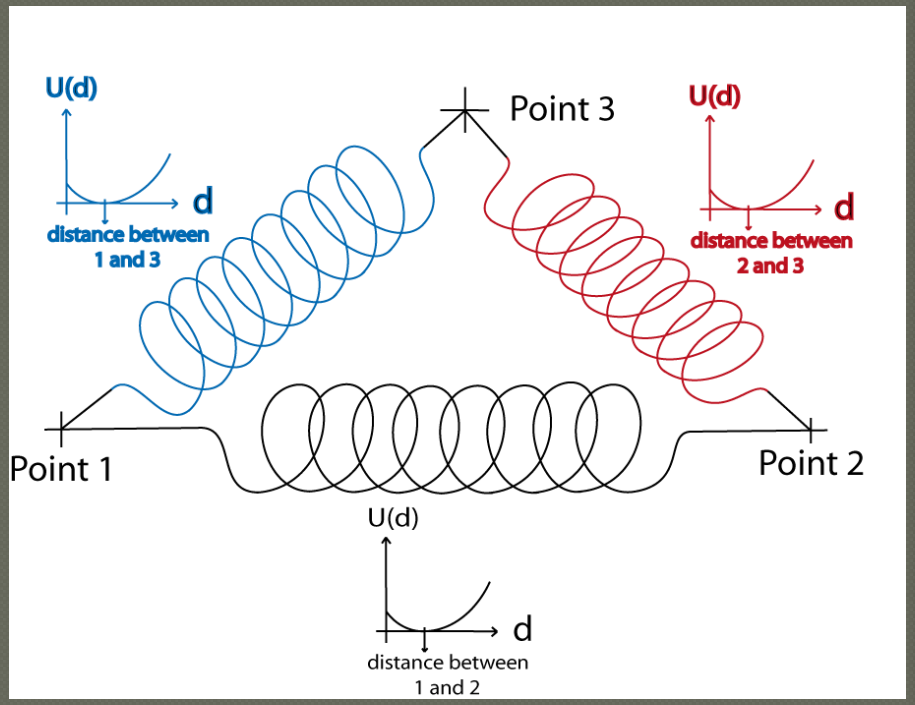
- Review of dimensionality reduction techniques
- SVD-MDS
- Pathway network analysis software
- 1st Application: Cerebral Malaria characterization
- 2nd Application: Treg characterization

Multidimensional Scaling initialized by Singular Value Decomposition



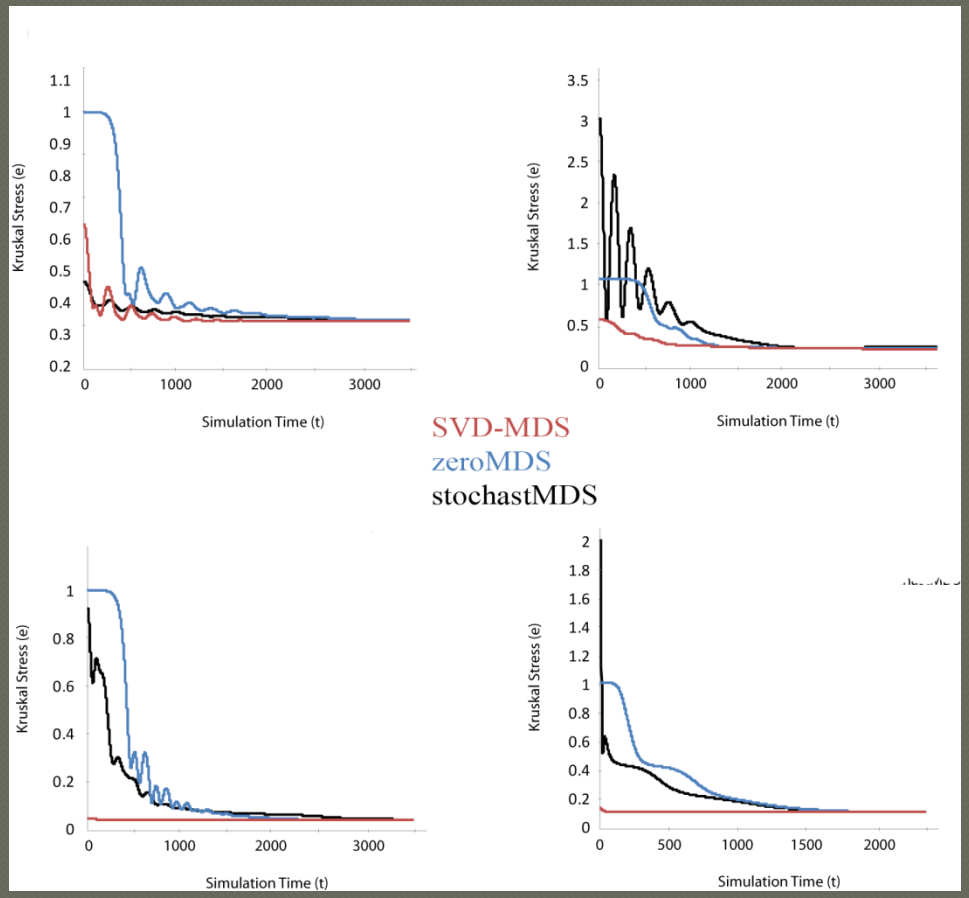
SVD

MDS using Spring analogy

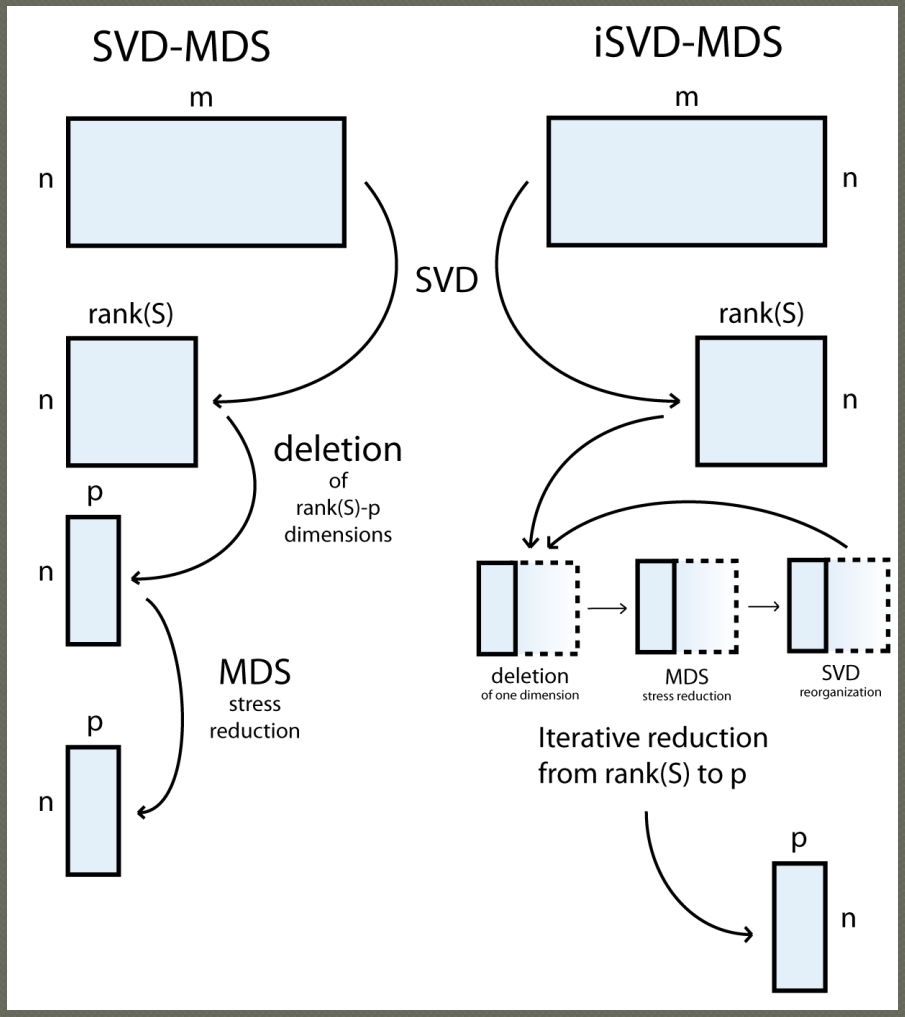


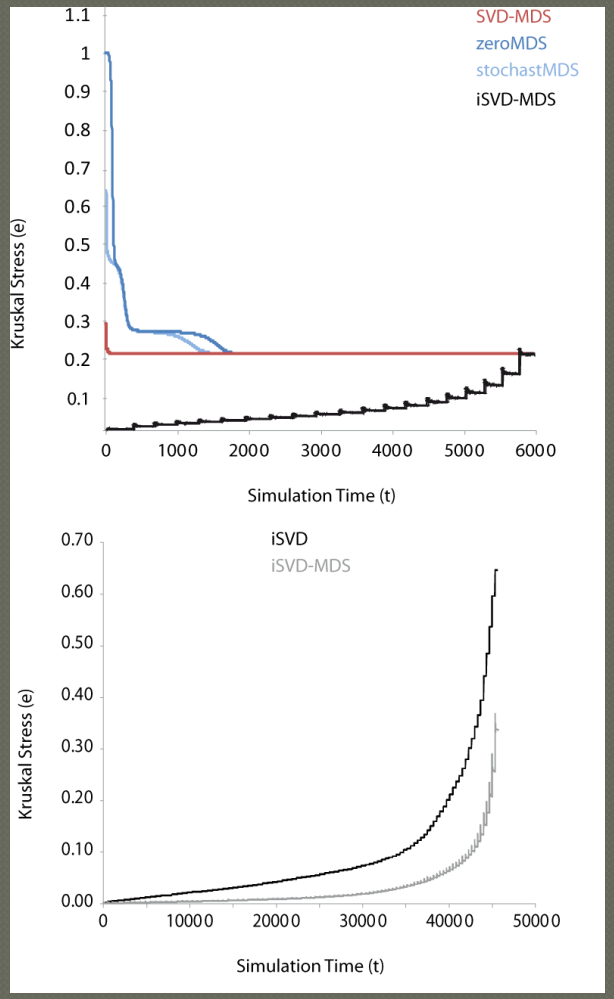
Comparison of different initialization strategies

ID	Dataset Name	No. of Instances	No. of Variables
d1	96Cell	96	32878
d2	96Cell_T	96	1553
d3	Iris	150	4
d4	Wine	178	13
d5	Stochast 200	200	50
d6	CCYier	516	12
d7	Pima	768	9
d8	96Cell_T transposed	1553	96
d9	Secom	1567	590
d10	Ozone	2565	72
d11	Stochast 3000	3000	300
d12	Ecoli	4288	7
d13	Wave	5000	22



Iterative SVD-MDS

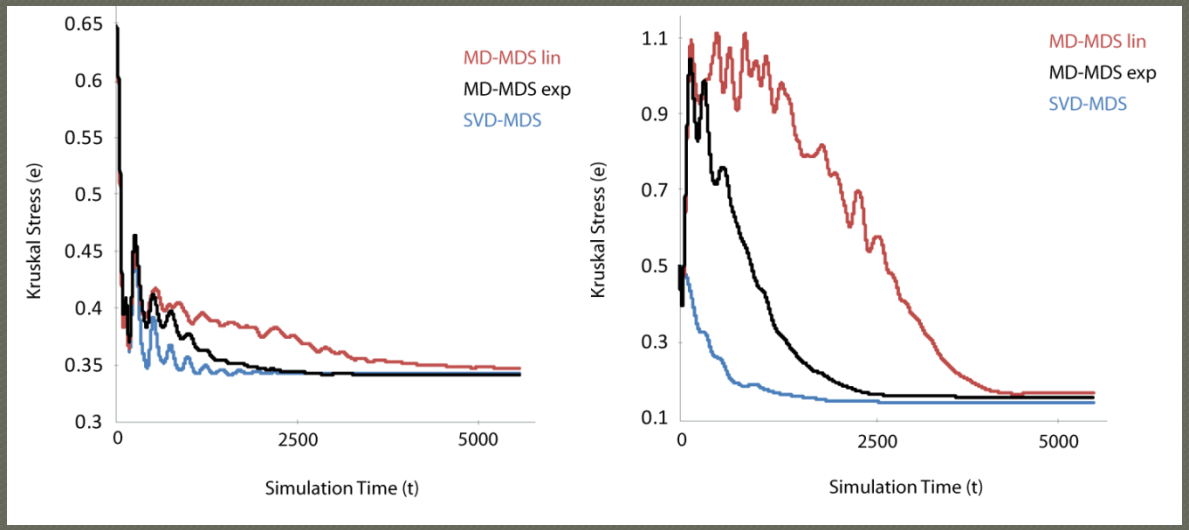


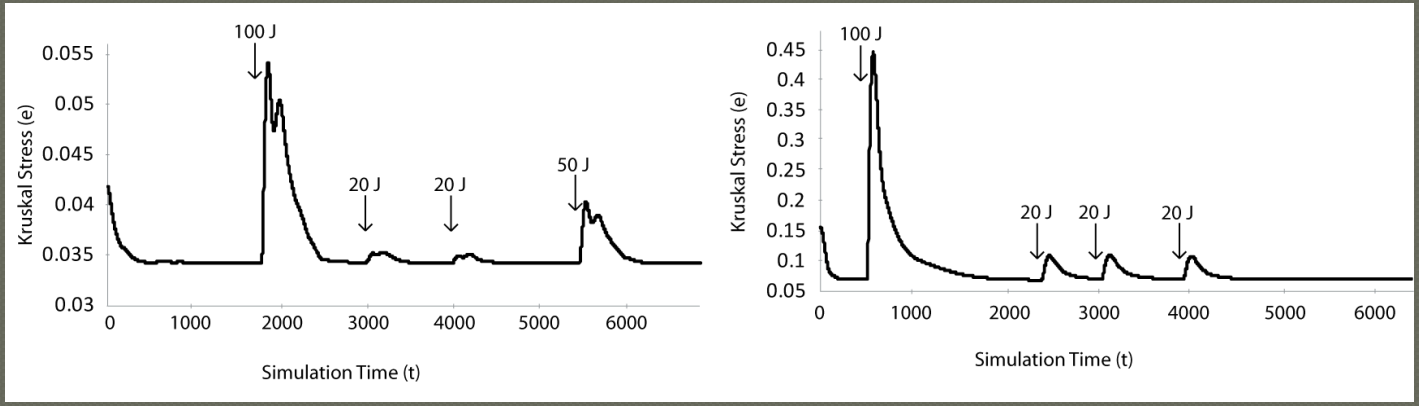


Adding stochasticity to avoid local minima ?

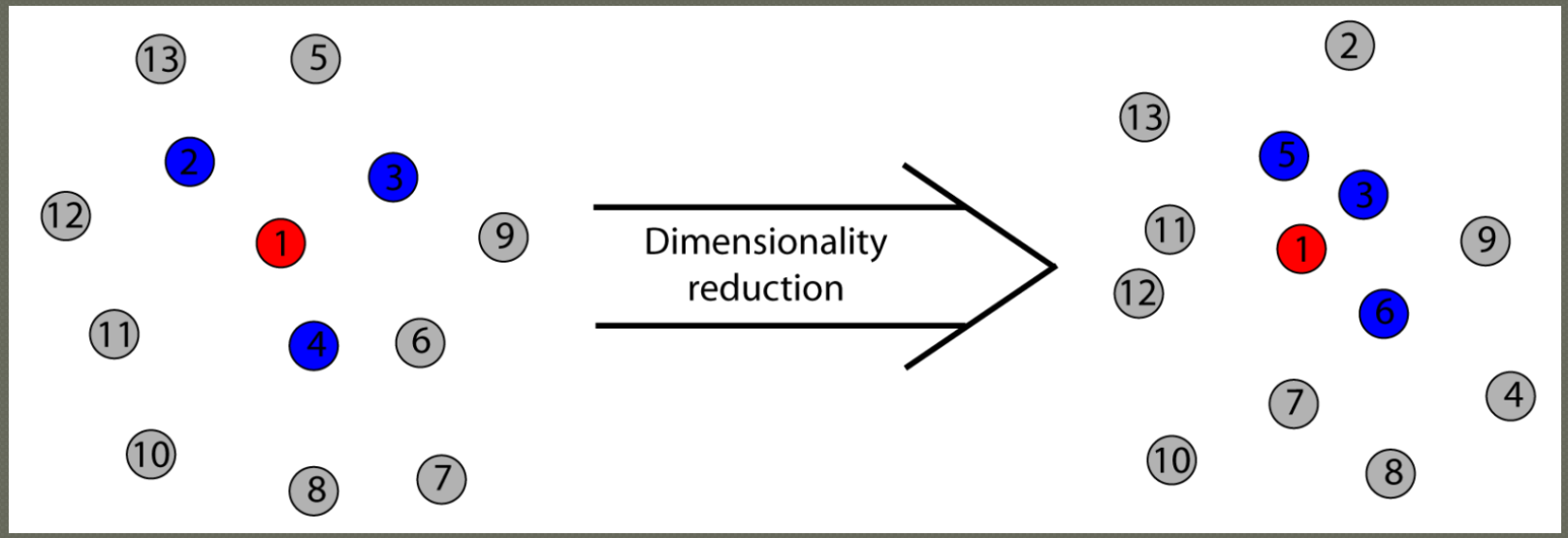
Stochastic force

$$F_{stoch}\left(x^i\right) = -T * s(t)$$





Entourage parameter
for evaluating local deformation



$$Ent_k = \frac{\sum_{i=1}^{n} G_i}{G}$$

with $G_i = card(N_i^{ref} \cap N_i^{new})$

and $G = nk$

SVD-MDS

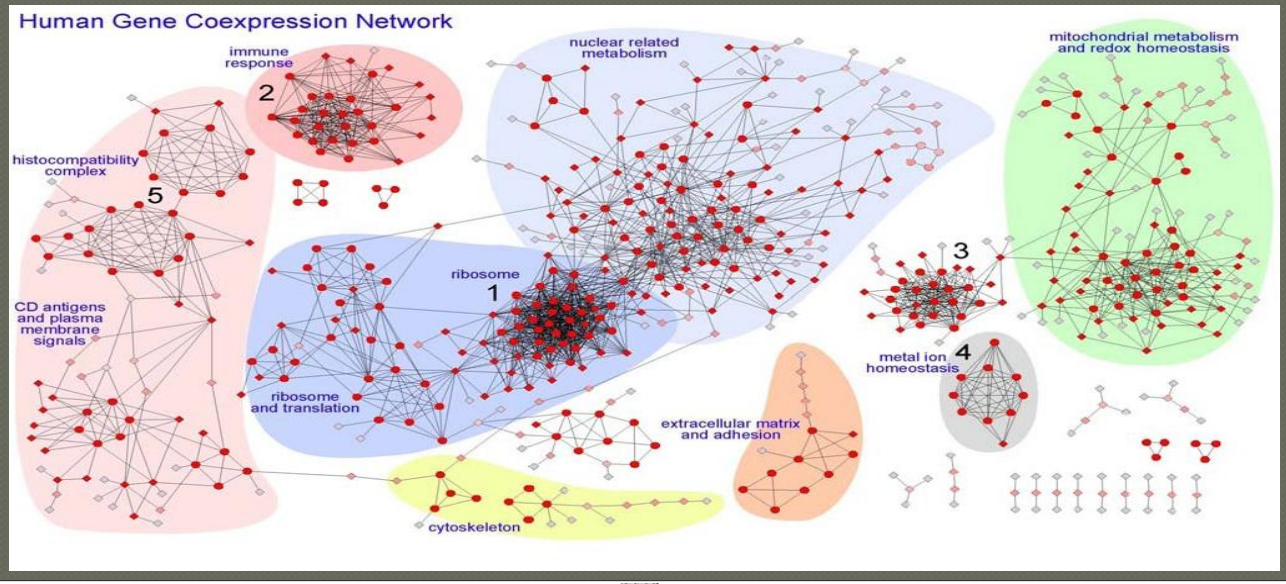
- The use of SVD is the best initialization strategy for MDS;
- Iterative dimensionality reduction does not improve the optimization;
- Adding stochastic also does not improve the process;
- Using parameter for local change assessment we prove that our representation is generally more accurate than the one obtain with our other MDS algorithm.

Bécavin et al., Bioinformatics, *Submitted*

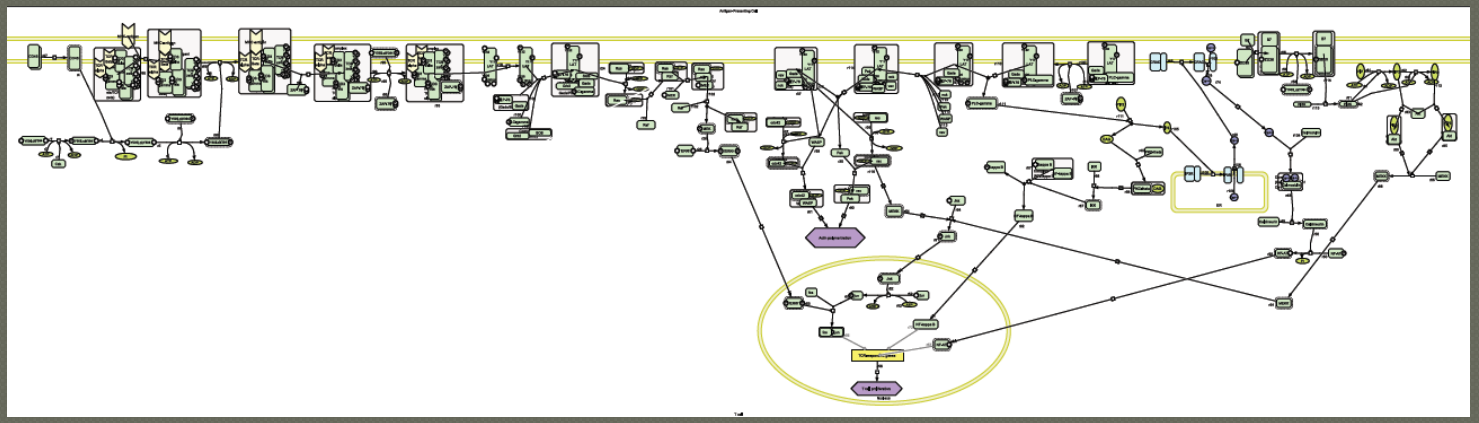
Index

- Review of dimensionality reduction techniques
- SVD-MDS
- Pathway network analysis software
- 1st Application: Cerebral Malaria characterization
- 2nd Application: Treg characterization

Gene expression and Pathway structure integration



Prieto et al. , Plos One 2008



Retrieving and curation of pathways:
List of species and list of reactions

XML file downloaded from PANTHER

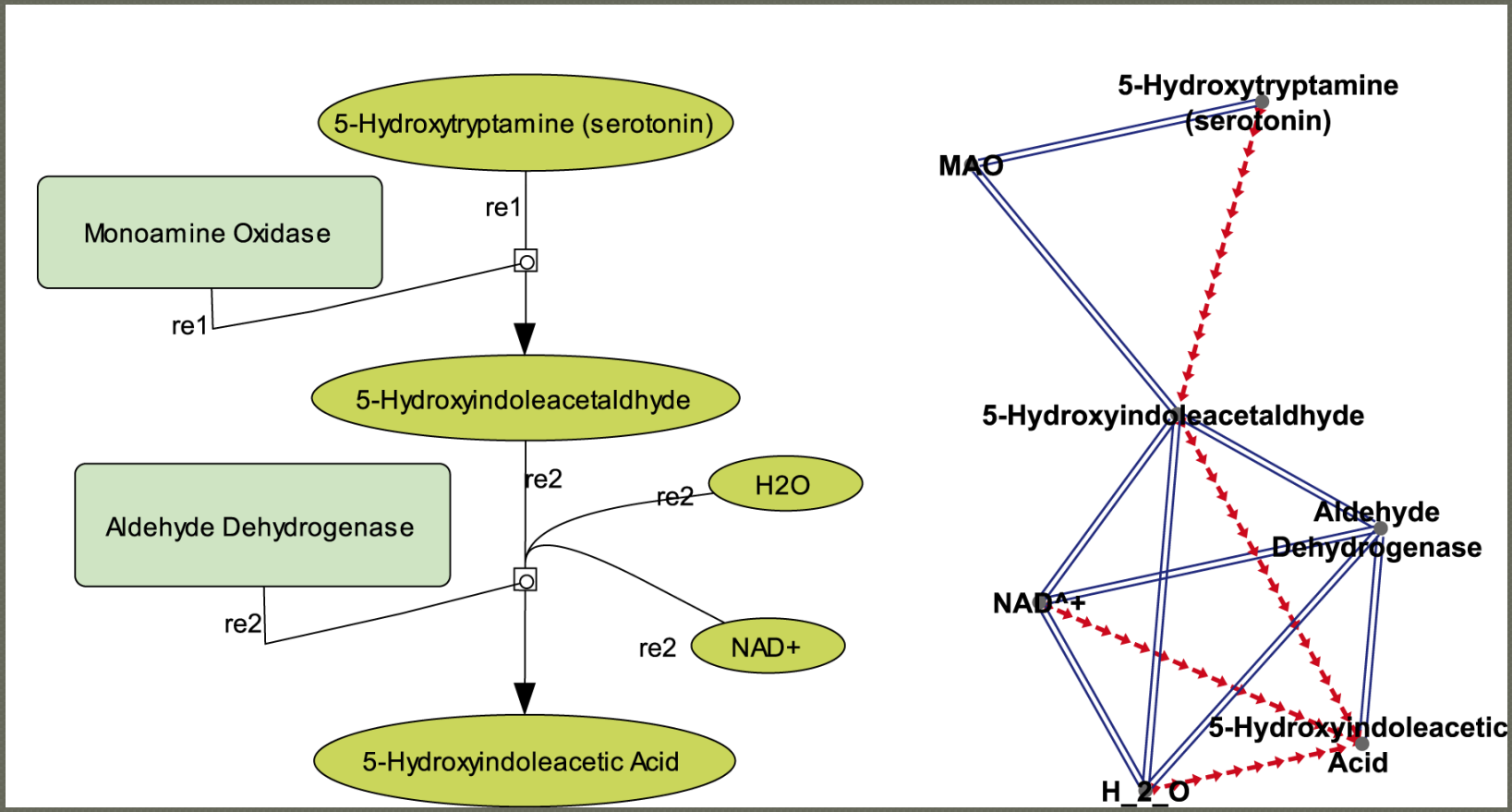
class	id	name	compartment	positionCompartment	included	LongName	Accession	Synonim	XMLname
SIMPLE_MOLECULE	s482	ADP	c5	inside		ADP	S01854	Synonym not specified	ADP
SIMPLE_MOLECULE	s305	ADP	c3	inside		ADP	S01854	Synonym not specified	ADP
PROTEIN	s425	AKT	c5	inside		Akt	P01332	PKB,Protein Kinase B	Akt
SIMPLE_MOLECULE	s481	ATP	c5	inside		ATP	S01851	Synonym not specified	ATP
SIMPLE_MOLECULE	s304	ATP	c3	inside		ATP	S01851	Synonym not specified	ATP
PHENOTYPE	s271	Actin polyme	c5	inside		Actin polyme	U02632	Synonym not specified	Actin polyme
PROTEIN	s337	B7	c2	transmembrane		B7	P01337	Synonym not specified	B7
PROTEIN	s320	CD28	c5	transmembrane		CD28	P01313	Synonym not specified	CD28
PROTEIN	s386	CD45	c5	transmembrane		CD45	P01329	Leukocyte common antig	CD45
PROTEIN	s19	CD45	c5	transmembrane		CD45	P01329	Leukocyte common antig	CD45
PROTEIN	s493	CRAC	c5	transmembrane		CRAC	P01328	Synonym not specified	CRAC
ION	s500	Ca^2+	default	inside		Ca++	U02631	Synonym not specified	Ca++
ION	s494	Ca^2+	c5	inside		Ca++	U02631	Synonym not specified	Ca++
ION	s445	Ca^2+	c4	inside		Ca++	U02631	Synonym not specified	Ca++
PROTEIN	s495	Calcineurin	c5	inside		Calcineurin	P01315	Synonym not specified	Calcineurin
PROTEIN	s294	Calmodulin	c5	inside		Calmodulin	P01305	Synonym not specified	Calmodulin
COMPLEX	s334	Complex;PIP	c5	inside	Akt;PIP2;	Long name n	U05532	Synonym not specified	Complex;Akt
COMPLEX	s335	Complex;PIP	c5	inside	Akt;PIP3;	Long name n	U05543	Synonym not specified	Complex;Akt
COMPLEX	s344	Complex;CD	c5	transmembrane	B7;CD28;	Long name n	U05549	Synonym not specified	Complex;B7;
COMPLEX	s343	Complex;CD	c5	transmembrane	B7;CD28;	Long name n	U05553	Synonym not specified	Complex;B7;

Species list for each pathway

type	id	reactants	products	modifiers	reversible
HETERODIMER_ASSOCIATION	r100	s406,s204,s411,s309,	s442,		false
HETERODIMER_ASSOCIATION	r101	s490,s415,	s449,		false
DISSOCIATION	r102	s449,	s247,s490,	s487,	false
TRANSPORT	r103	s500,	s494,	s493,	false
HETERODIMER_ASSOCIATION	r104	s494,s294,	s451,		false
STATE_TRANSITION	r105	s452,	s452,	s281,	false
STATE_TRANSITION	r106	s445,	s494,	s452,	false
STATE_TRANSITION	r107	s464,s481,	s461,s482,	s486,	false
DISSOCIATION	r110	s461,	s419,s455,		false
DISSOCIATION	r111	s492,	s282,s281,	s455,	false
HETERODIMER_ASSOCIATION	r112	s425,s330,	s335,		false
HETERODIMER_ASSOCIATION	r113	s419,s408,s485,s484,s4	s471,		false
DISSOCIATION	r114	s471,	s477,s485,s484,		false
STATE_TRANSITION	r115	s498,	s498,	s344,	false
KNOWN_TRANSITION_OMITTE	r116	s491,	s496,		false
STATE_TRANSITION	r16	s4,	s406,	s437,	false
STATE_TRANSITION	r24	s242,s488,	s243,s489,	s442,	false
HETERODIMER_ASSOCIATION	r25	s243,s487,	s490,		false

Reactions list for each pathway

Obtaining the list of reactions



Association of all species to their probes

67% (908/1345) of proteins associated

Probes list from panther

Probe
108931
111111
108546
107975
107703
107477
106774
105950
105711
111356
104619
104387
103919
103666
102968
102762
102438
111840

Species list from panther

class	id	name
SIMPLE_MOLECULE	s482	ADP
SIMPLE_MOLECULE	s305	ADP
PROTEIN	s425	AKT
SIMPLE_MOLECULE	s481	ATP
SIMPLE_MOLECULE	s304	ATP
PHENOTYPE	s271	Actin polyme
PROTEIN	s337	B7
PROTEIN	s320	CD28
PROTEIN	s386	CD45
PROTEIN	s19	CD45
PROTEIN	s493	CRAC
ION	s500	Ca^2+
ION	s494	Ca^2+
ION	s445	Ca^2+
PROTEIN	s495	Calcineurin
PROTEIN	s294	Calmodulin
COMPLEX	s334	Complex;PIP
COMPLEX	s335	Complex;PIP

Probes versus species association

Probe_ID	Species associated	Gene_Symbol	Gene_Name	Panther_Family_Name	Panther_Subfamily_Name	Panther_Function	Panther_Process
248820	Actin	ACTC	actin, alpha, ca	ACTIN	null	Actin and actin rela	Cell structure|Exo
696998	Actin	LOC440888	null	ACTIN	ACTIN-LIKE PROTEIN 3, ACL3	Actin and actin rela	Cell structure|Mit
10693438	Actin	ACTR2	ARP2 actin-rela	ACTIN	ACTIN-RELATED PROTEIN 2,	Actin and actin rela	Cell structure|Mit
199711	ADAM10	ADAM10	ADAM metallo	ADAM (A DISINTEGRIN A	ADAM 10	Metalloprotease|P	Proteolysis;Signal
212988	ADAM10	ADAM10	ADAM metallo	ADAM (A DISINTEGRIN A	ADAM 10	Metalloprotease|P	Proteolysis;Signal
141207		SLC25A6	solute carrier f	MITOCHONDRIAL CARRI	ADP,ATP CARRIER PROTEIN	Transporter;Transfe	Nucleoside, nucle
226681		SLC25A31	solute carrier f	MITOCHONDRIAL CARRI	ADP,ATP CARRIER PROTEIN	Transporter;Transfe	Nucleoside, nucle
113928	Adenine phosphoribosyltransferas	APRT	adenine phosp	null	null	null	null
169183	Adenomatous Polyposis Coli	APC	adenomatosis	ADENOMATOUS POLYPC	ADENOMATOUS POLYPOSIS	Microtubule family	Intracellular signal
173633	Adenomatous Polyposis Coli	APC2	adenomatosis	ADENOMATOUS POLYPC	ADENOMATOUS POLYPOSIS	Microtubule family	Intracellular signal
100892		ADK	adenosine kina	SUGAR KINASE RELATED	ADENOSINE KINASE	Kinase|Nucleotide	Nucleoside, nucle
119977	Adenosine deaminase	ADA	adenosine dea	ADENOSINE DEAMINASE	ADENOSINE DEAMINASE	Deaminase|Hydrol	Nucleoside, nucle
130012	Adenosine deaminase	ADAL	adenosine dea	ADENOSINE DEAMINASE	ADENOSINE DEAMINASE-RE	Deaminase|Hydrol	Nucleoside, nucle
162894	Adenosine deaminase	ADAL	adenosine dea	ADENOSINE DEAMINASE	ADENOSINE DEAMINASE-RE	Deaminase|Hydrol	Nucleoside, nucle
183103	Adenosine deaminase	CECR1	cat eye syndror	ADENOSINE DEAMINASE	CAT EYE SYNDROME CRITICA	Deaminase|Signali	Developmental pr
104782		ADSL	adenylosuccina	ASPARTATEAMMONIA/A	ADENYLOSUCCINATE LYASE	Other lyase|Lyase	Nucleoside, nucle
107396	adenylate cyclase	ADCY8	adenylate cycla	ADENYLATE AND GUANY	ADENYLATE CYCLASE TYPE V	Lyase|Cyclase|Ade	Nucleoside, nucle
110855	adenylate cyclase	ADCY2	adenylate cycla	ADENYLATE AND GUANY	ADENYLATE CYCLASE TYPE II	Lyase|Cyclase|Ade	Nucleoside, nucle
122069	adenylate cyclase	ADCY3	adenylate cycla	ADENYLATE AND GUANY	ADENYLATE CYCLASE TYPE II	Lyase|Cyclase|Ade	Nucleoside, nucle
125794	adenylate cyclase	TSCOT	thymic stromal	ADENYLATE CYCLASE	DRUG EFFLUX PROTEIN-REL	Molecular function	Transport|Small m

Get Panther probes association

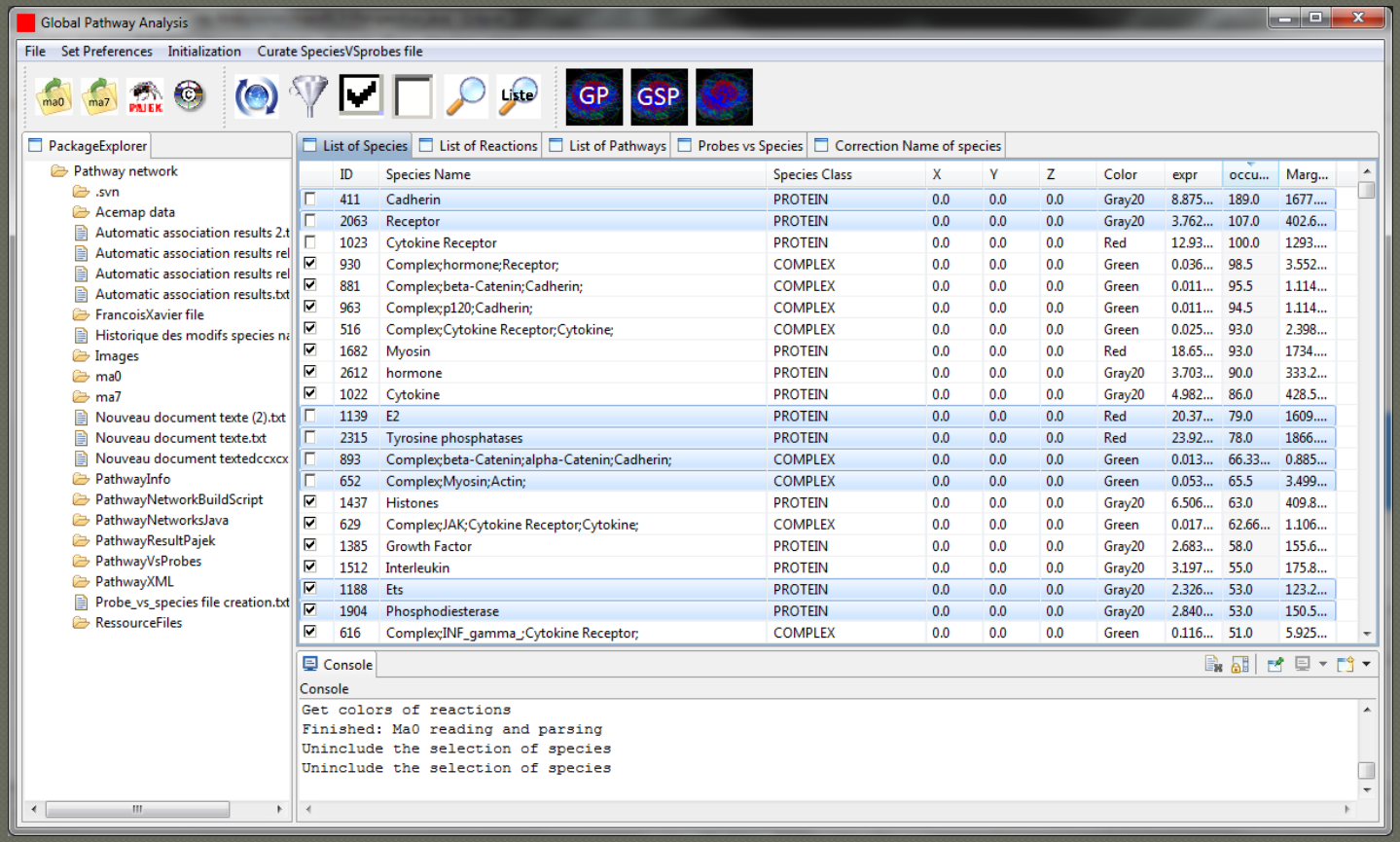
Get already associated probes

Get species not associated from this pathway

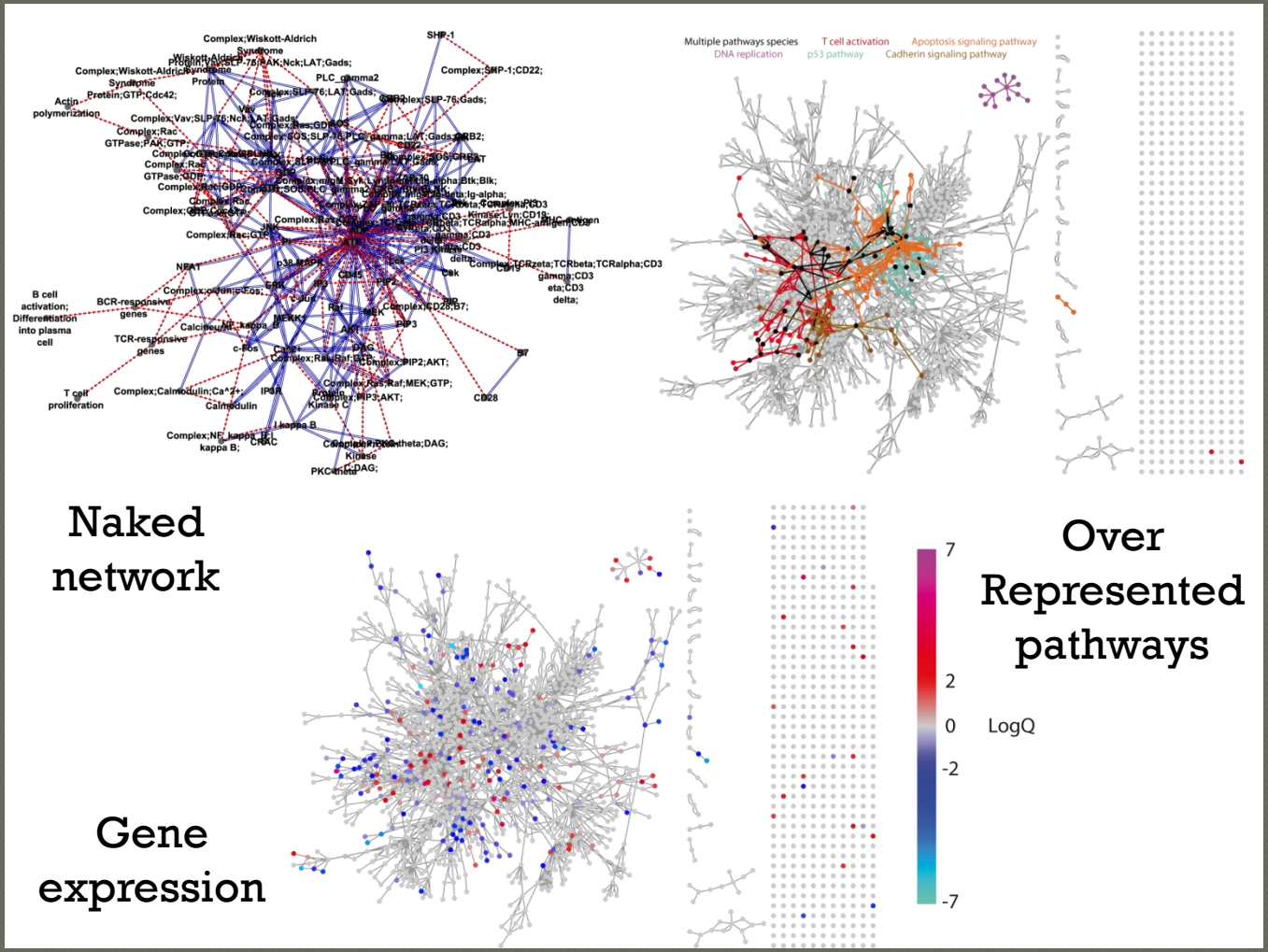
Probe Panther	Probe PathwayAnalysis	Species PathwayAnalysis	Species associated	LongName	Gene_Symbol	Gene_Name	Panther_Family_Name	Panther_Subfamily_Name	Panther_Function	Panther_Process
101444			NF_kappa_B	Nuclear fact	NFKB2	nuclear facto	ANKYRIN REPEAT-CON	NUCLEAR FACTOR NF-KAPI	Other transcriptio	Oncogene|Intrac
102071	102071	TCRbeta	TCRbeta		null	null	T-CELL RECEPTOR BETA	T-CELL RECEPTOR BETA CH.	Immunoglobulin r	Immunity and de
102086	null	null	null		null	null	null	null	null	null
102438	102438	Vav	Vav		VAV1	vav 1 oncoge	DBL RELATED	VAV1	G-protein modula	Calcium mediate
102762	null	null	null		LOC441971	null	PHOSPHATIDYLINOSIT	PHOSPHATIDYLINOSITOL 3	Kinase|Other kina	Intracellular sigr
102968	102968	Ras	Ras	Rat Sarcoma	KRAS	v-Ki-ras2 Kirs	RAS-RELATED GTPASE	RAS	G-protein|Small G	Oncogene|Intrac
103666	103666	Ras	Ras	Rat Sarcoma	NRAS	neuroblastor	RAS-RELATED GTPASE	RAS	G-protein|Small G	Oncogene|Intrac
103919	null	null	null		PIK3CD	phosphoinos	PHOSPHATIDYLINOSIT	PHOSPHATIDYLINOSITOL 3	Kinase|Other kina	Intracellular sigr
104387	104387	PAK	PAK	P21 activate	PAK2	p21 (CDKN1A)	MAPKK-RELATED SERIN	P21-ACTIVATED KINASE, PA	Kinase|Protein kir	Protein metaboli
104619			Rac	Ced-10@Rac	null	null	RAS-RELATED GTPASE	RAC GTPASE	G-protein|Small G	Cell surface rece
105711	105711	JNK	JNK	C-Jun N-term	MAPK8	mitogen-acti	CDC2-RELATED KINASE	JNK	Protein kinase|Kir	MAPKKK cascade|
105950	null	null	null		GRAP2	GRB2-relatec	SH2 DOMAIN ADAPTOI	GRB2-RELATED ADAPTOR I	Transmembrane re	Cell surface rece
106774			Rac	Ced-10@Rac	RAC1	ras-related C	RAS-RELATED GTPASE	RAC GTPASE	G-protein|Small G	Cell surface rece
107477	null	null	null		ZDHHC16	zinc finger, D	ZINC FINGER DHHC DC	null	Miscellaneous fur	Biological proces
107703	107703	CD28	CD28		CD28	CD28 antigen	CYTOTOXIC T-LYMPHO	T-CELL-SPECIFIC SURFACE	Immunoglobulin r	Immunity and de
107975			Tyrosine phosphata	Protein-tyros	PTPRC	protein tyros	PROTEIN-TYROSINE PH	RECEPTOR PROTEIN TYROS	Protein phosphata	Cell communicat
108546	null	null	null		null	null	null	null	null	null
108931	null	null	null		CD3Z	CD3Z antigen	T-CELL SURFACE GLYCC	null	Receptor;Defense,	Intracellular sigr
111111	null	null	null		ITPR1	inositol 1,4,5	RYANODINE RECEPTOR	INOSITOL 1,4,5-TRISPHOSI	Other ligand-gate	Calcium mediate
111356			NF_kappa_B	Nuclear fact	NFKBIA	nuclear facto	ANKYRIN REPEAT-CON	NF KAPPA B INHIBITOR I-K	Select regulatory n	Nucleoside, nucl
111840	null	null	null		C10orf75	chromosome	null	null	null	null
112906	112906	TCRbeta	TCRbeta		null	null	T-CELL RECEPTOR BETA	T-CELL RECEPTOR BETA CH.	Immunoglobulin r	Immunity and de
112923			MKK2	mitogen-act	MAP2K2	mitogen-acti	MAPKK-RELATED SERIN	MITOGEN ACTIVATED PRO	Kinase|Protein kir	Protein metaboli
112952	null	null	null		PDZK6	PDZ domain	FAMILY NOT NAMED	SUBFAMILY NOT NAMED	Molecular functio	Biological proce

Different types of networks that can be constructed

Selection of species and reactions ⟶ Nodes and edges of the network



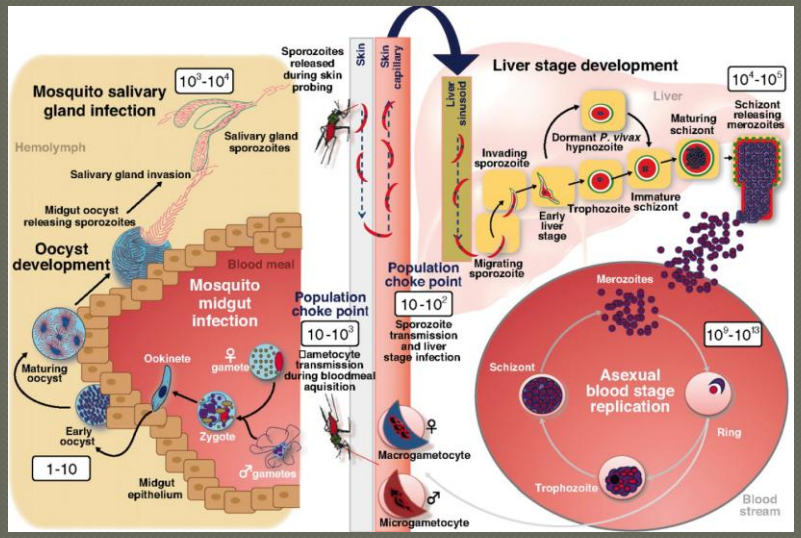
Different types of networks that can be constructed



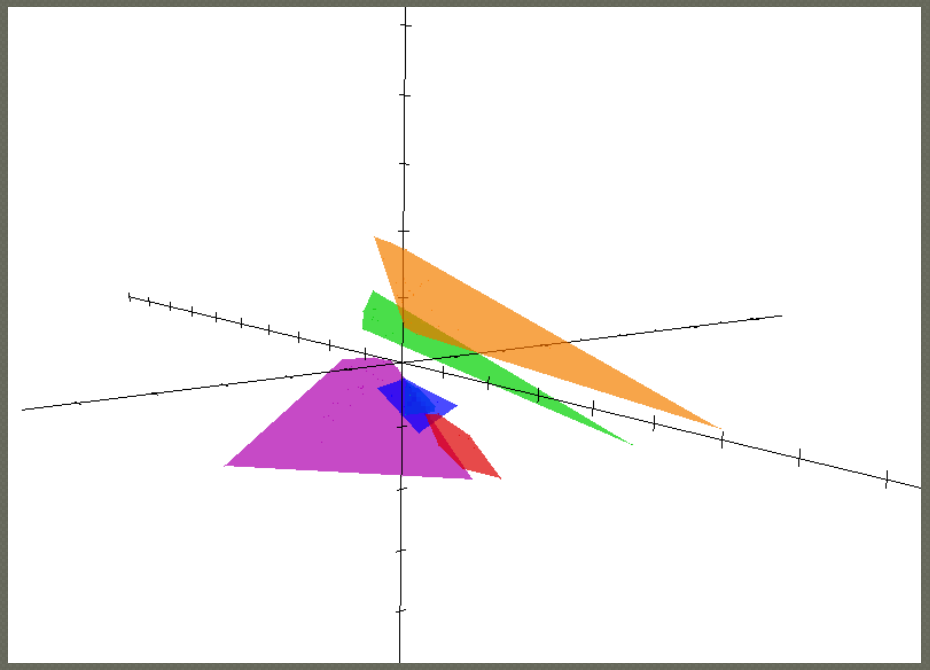
Index

- Review of dimensionality reduction techniques
- SVD-MDS
- Pathway network analysis software
- 1st Application: Cerebral Malaria characterization
- 2nd Application: Treg characterization

Cerebral Malaria characterization using auto-antibodies and Cytokines



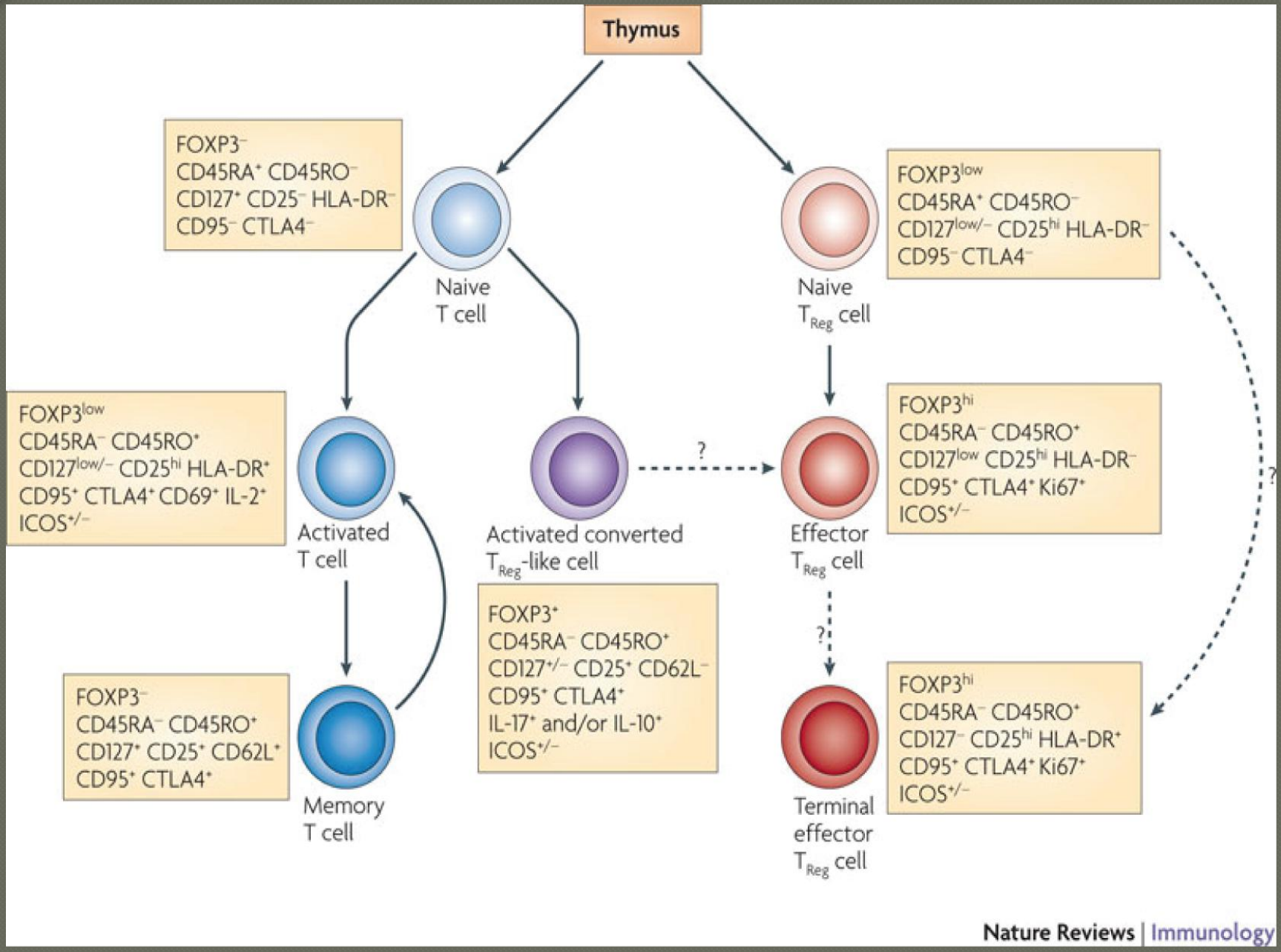
IgG Autoantibody to Brain Beta Tubulin III Associated with Cytokine Cluster-II Discriminate Cerebral Malaria in Central India. D. Bansal, F. Herbert, P. Lim1, P. Deshpande, C. Bécavin,… …, A. Benecke, S.Pied., PLoS ONE, 2009



Cerebral malaria (CM)
Severe non-cerebral malaria (SM)
Mild malaria (MM)
Recovered cerebral malaria (ex-CM)
Endemic controls (EC)
Non-endemic controls (NEC)

Index

- Review of dimensionality reduction techniques
- SVD-MDS
- Pathway network analysis software
- 1st Application: Cerebral Malaria characterization
- 2nd Application: Treg characterization


Thymus
FOXP3⁻
CD45RA⁺ CD45RO⁻
CD127⁺ CD25⁻ HLA-DR⁻
CD95⁻ CTLA4⁻
Naive
T cell
Naive
T$_{Reg}$ cell
FOXP3^{low}
CD45RA⁺ CD45RO⁻
CD127$^{low/-}$ CD25^{hi} HLA-DR⁻
CD95⁻ CTLA4⁻
FOXP3^{low}
CD45RA⁻ CD45RO⁺
CD127$^{low/-}$ CD25^{hi} HLA-DR⁺
CD95⁺ CTLA4⁺ CD69⁺ IL-2⁺
ICOS$^{+/-}$
Activated
T cell
Activated converted
T$_{Reg}$-like cell
?
Effector
T$_{Reg}$ cell
FOXP3^{hi}
CD45RA⁻ CD45RO⁺
CD127^{low} CD25^{hi} HLA-DR⁻
CD95⁺ CTLA4⁺ Ki67⁺
ICOS$^{+/-}$
FOXP3⁺
CD45RA⁻ CD45RO⁺
CD127$^{+/-}$ CD25⁺ CD62L⁻
CD95⁺ CTLA4⁺
IL-17⁺ and/or IL-10⁺
ICOS$^{+/-}$
FOXP3⁻
CD45RA⁻ CD45RO⁺
CD127⁺ CD25⁺ CD62L⁺
CD95⁺ CTLA4⁺
Memory
T cell
?
Terminal
effector
T$_{Reg}$ cell
FOXP3^{hi}
CD45RA⁻ CD45RO⁺
CD127⁻ CD25^{hi} HLA-DR⁺
CD95⁺ CTLA4⁺ Ki67⁺
ICOS$^{+/-}$
?
Nature Reviews | Immunology

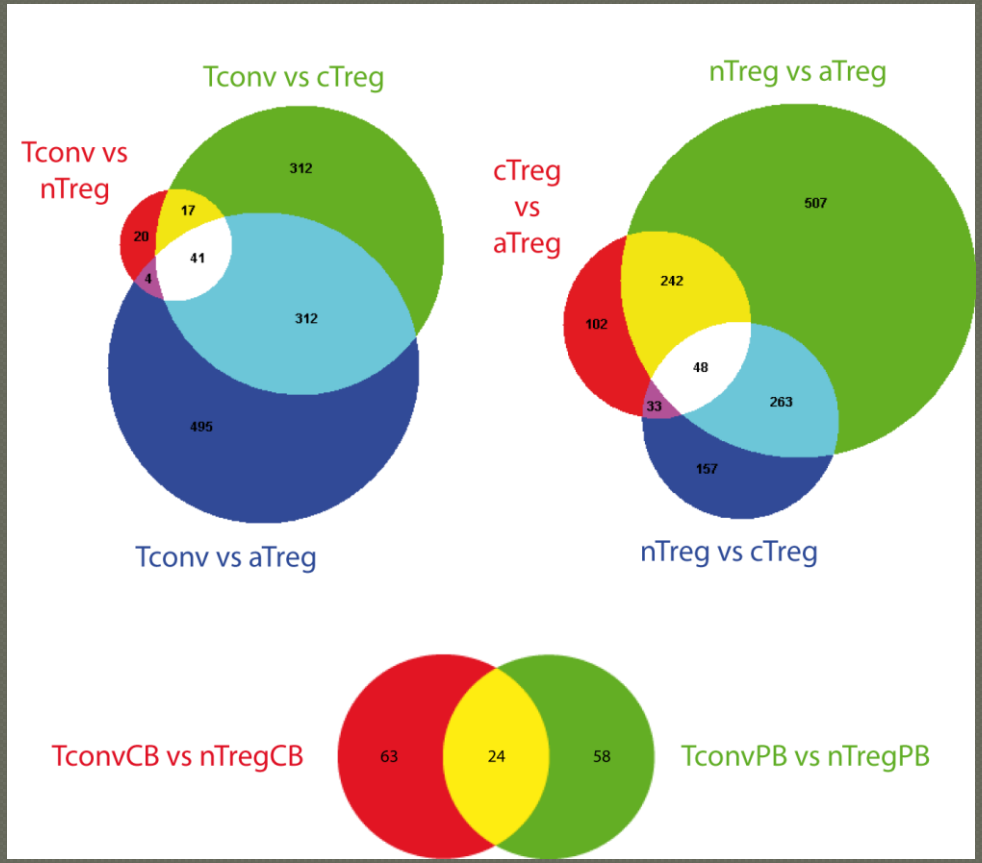
We have defined a multitude of gene sets (signatures) specific to each biological condition

Peripheral blood:

- aTreg (activated Treg)
CD4+CD25hiCD127-CD45RA-HLADR+
- cTreg (cytokine emiter Treg)
CD4+CD25hiCD127-CD45RA-HLADR-
- nTreg (naive Treg)
CD4+CD25hiCD127-CD45RA+HLADR-
- Tconv
CD4+CD25-CD45RA+HLADR-

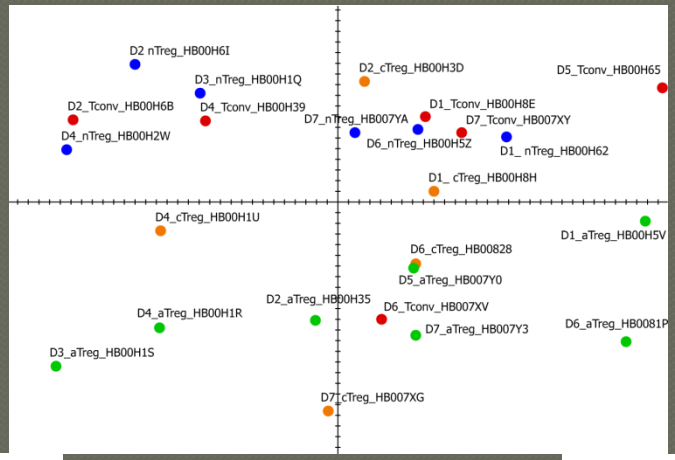
Cord Blood:

- nTreg (naive Treg)
CD4+CD25hiCD127-CD45RA+HLADR-
- Tconv
CD4+CD25-CD45RA+HLADR-

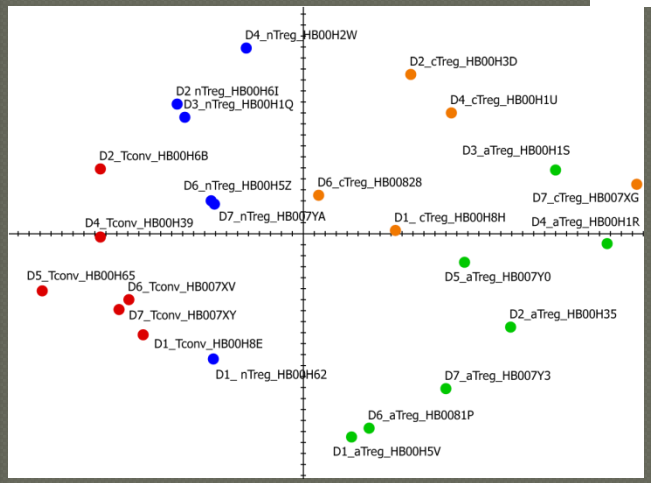


Representation (in correspondence basis) of the different Treg subsets for the different signatures constructed

Total signature
1201 probes over-represented and under-represented (in comparison to Tconv) for at least one Treg subset

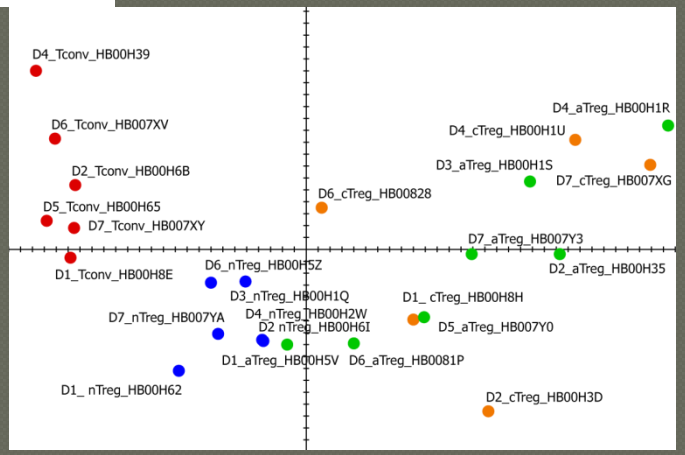


Core signature
41 probes over-represented and under-represented (in comparison to Tconv) for every Treg subset

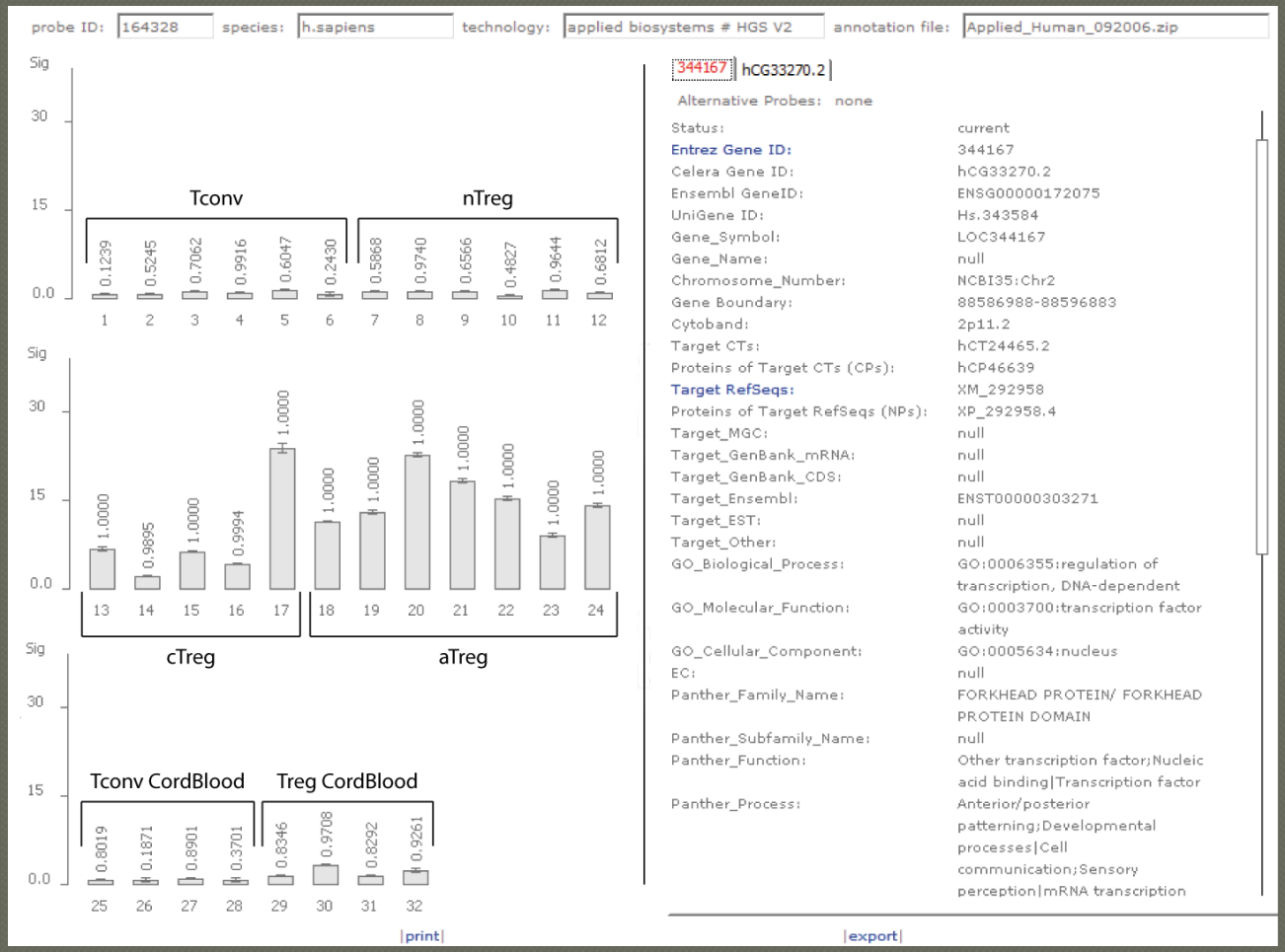


All probes

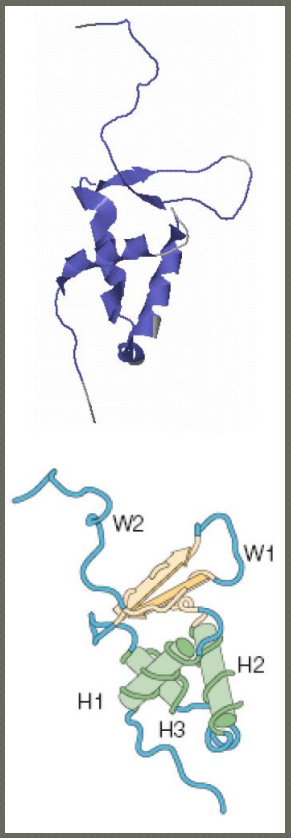
Tconv nTreg
cTreg aTreg



FOXLF a new transcription factor !

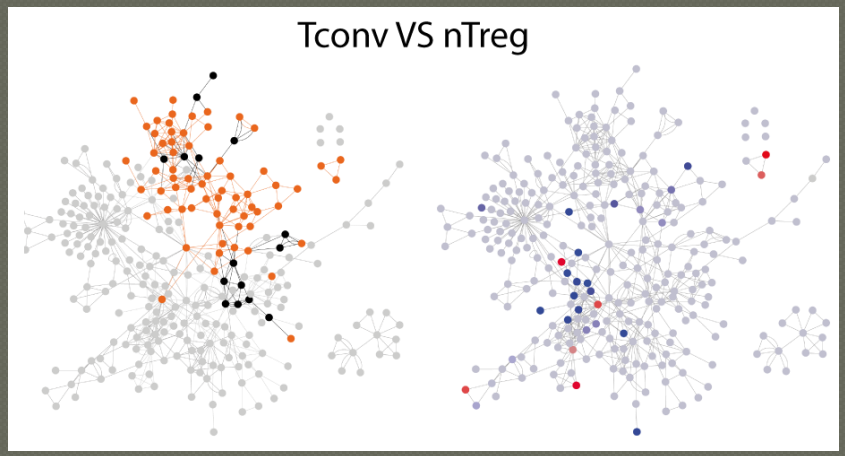


FOXLF



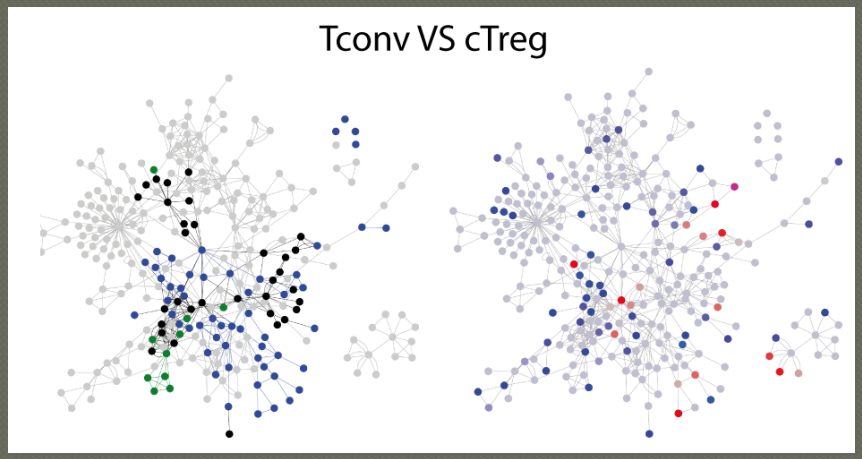
FOXP3

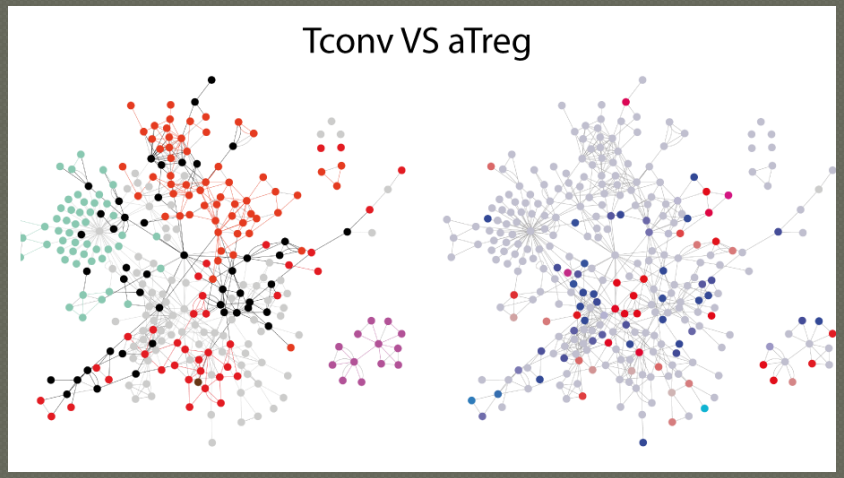
Search of the most representative pathways



Multiple pathways species
T cell activation
Apoptosis signaling pathway
DNA replication
p53 pathway
Cadherin signaling pathway
Interleukin signaling pathway
Inflammation mediated by chemokine and cytokine signaling pathway

Log of fold-change





Conclusion-perspective

- SVD-MDS: New MDS algorithm which is accurate and computationally efficient.

  New dimensionality reduction methods for the representation of high dimensional "omics" data. *C. Becavin, A. Benecke. Expert Review of Molecular Diagnosis*, will appear January 2011

  Molecular dynamics multidimensional scaling initialized by singular value decomposition leads to computationally efficient analysis of high dimensional data. C*. Becavin, N. Tchitchek, C. Mintsa-Eye, A. Lesne, A. Benecke submitted to Bioinformatics*

- Software developed for gene expression and Pathway structure integration

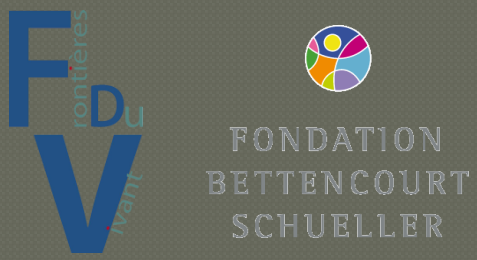
- Three different applications to biology

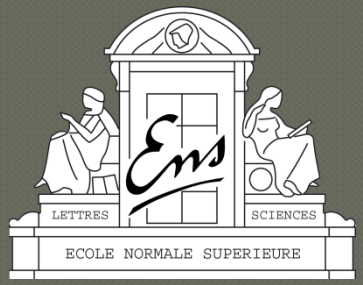
  IgG autoantibody to Brain beta Tubulin III associated with cytokine Cluster-II Discriminate Cerebral Malaria in Central India. *Bansal D, Herbert F, Lim P, Deshpande P, Becavin C, Guiyedi V, de Maria I, Rousselle JC, Namane A, Jain R, Cazenave PA, Mishra GC, Ferlini C, Fesel C, Benecke A, Pied S. Plos One 4(12) (2009)*

  HMGA1-dependent and independent 7SK RNA gene regulatory activity. Sebastian Eilebrecht, Christophe Becavin, Helene Leger, Bernd-Joachim Benecke, and Arndt Benecke. RNA Biology, In press.

  Publication of the results of the study on Treg subsets with Lars Rogge..

Remerciements





- All jury members: Alain Arneodo, Lars Rogge, Sylviane Pied, Jean Marc Victor, Andrei Zinovyev.

- My Thesis supervisors: Annick Lesne, Arndt Benecke.

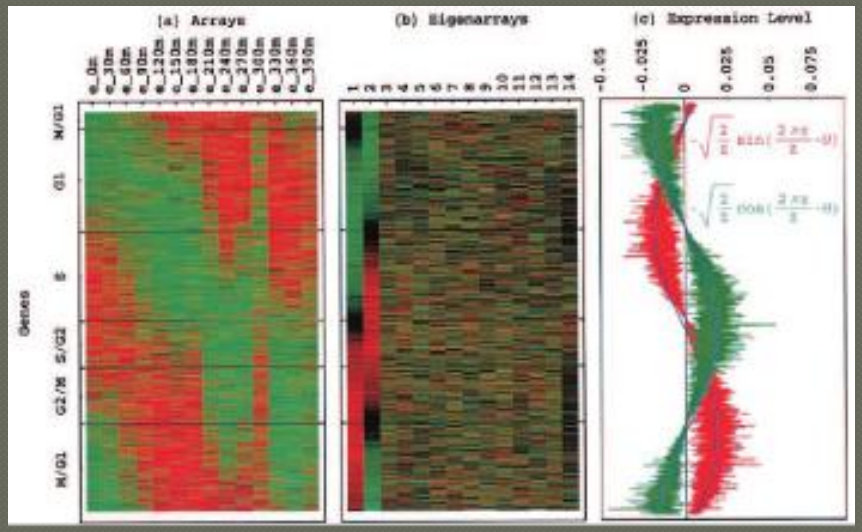
- All my colleagues: Nicolas, Felipe, Brice, Guillaume, Hélène, Sebastian, François-Xavier.

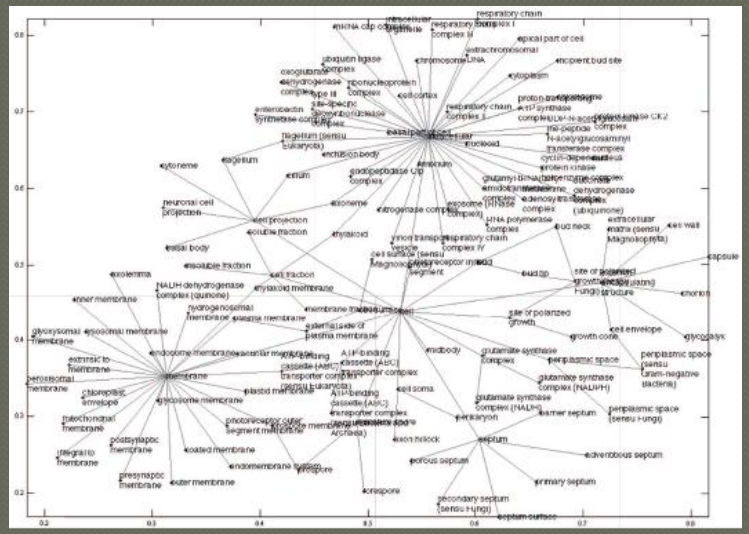
- My collaborators: Sylvia Maiella and the members of Lars Rogge's and Sylviane Pied's laboratories.

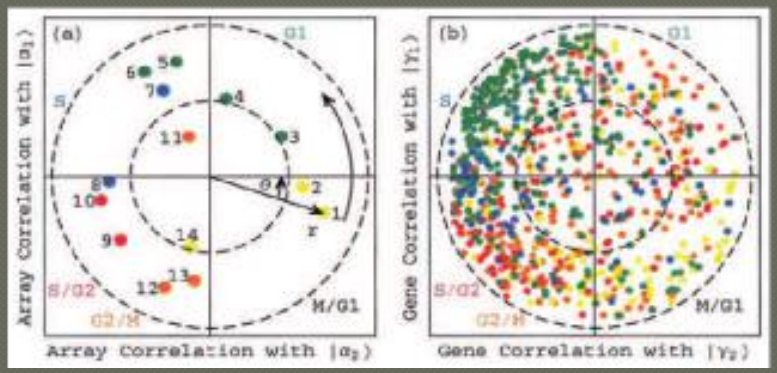
- All the people from my doctoral school.

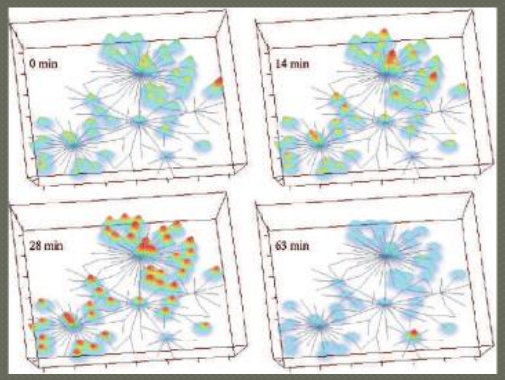
- Ma famille, mes amis.

Example of applications of dimensionality reduction techniques



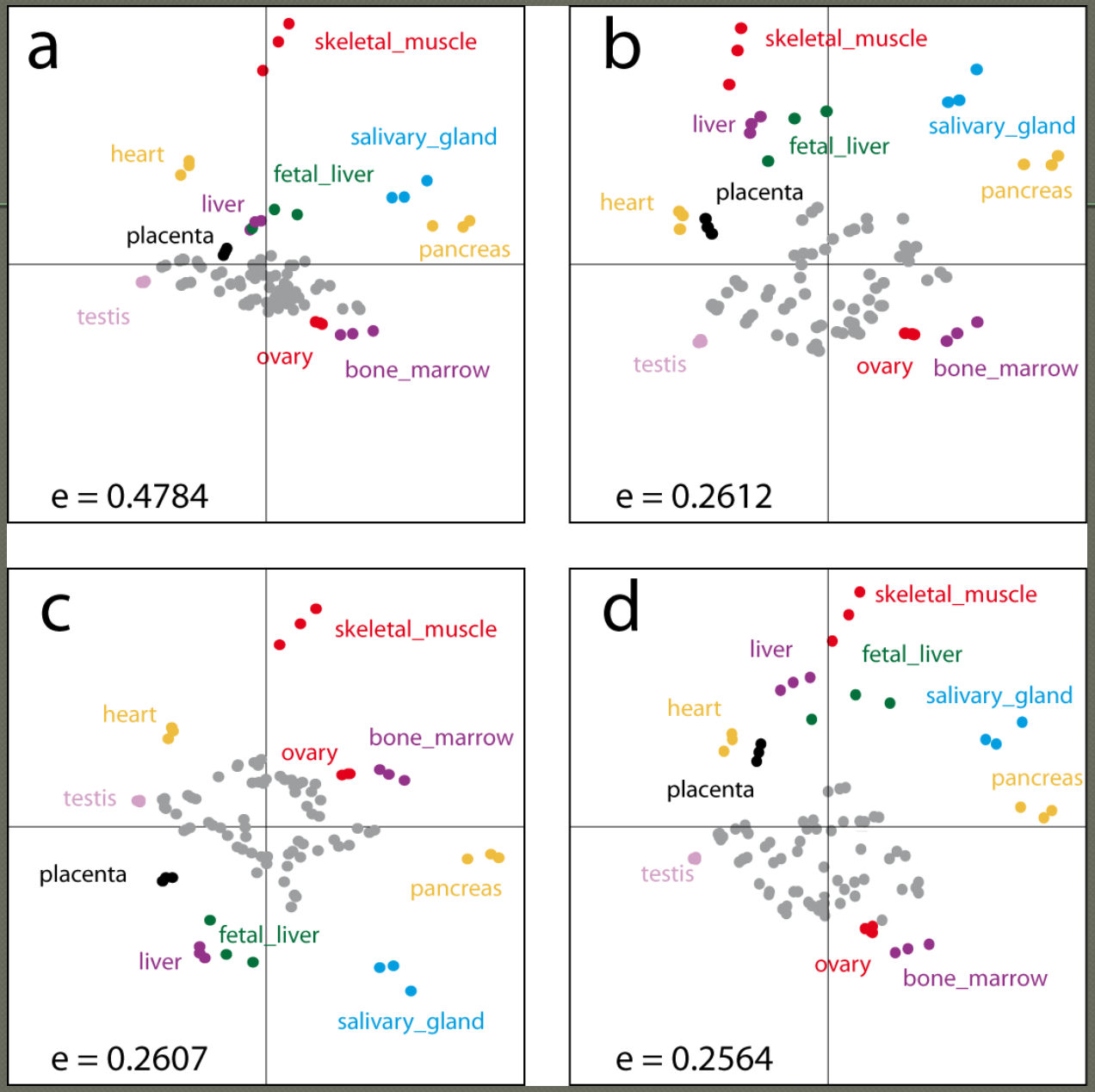






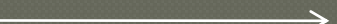
Alter et al., PNAS, 2000

Ebbels, Bioinformatics, 2006


a
skeletal_muscle
salivary_gland
heart
fetal_liver
liver
placenta
pancreas
testis
ovary
bone_marrow
e = 0.4784
b
skeletal_muscle
liver
fetal_liver
salivary_gland
pancreas
heart
placenta
testis
ovary
bone_marrow
e = 0.2612
c
skeletal_muscle
heart
bone_marrow
ovary
testis
placenta
pancreas
fetal_liver
liver
salivary_gland
e = 0.2607
d
skeletal_muscle
liver
fetal_liver
heart
salivary_gland
pancreas
placenta
testis
ovary
bone_marrow
e = 0.2564

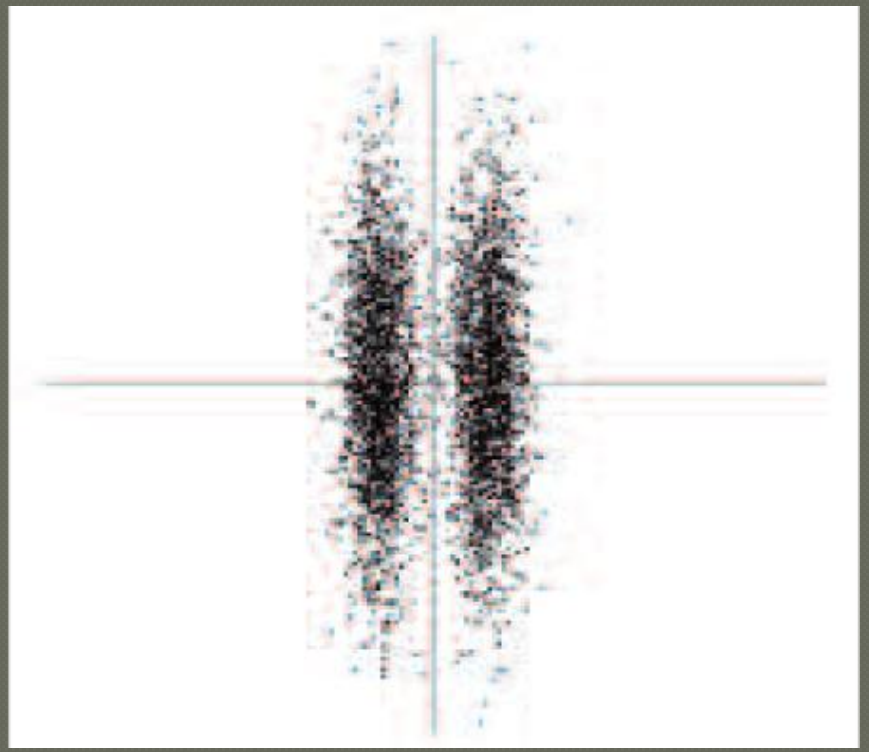
General view on linear analysis and new developments

Orthogonality
(e.g.: PCA, Factor Analysis)



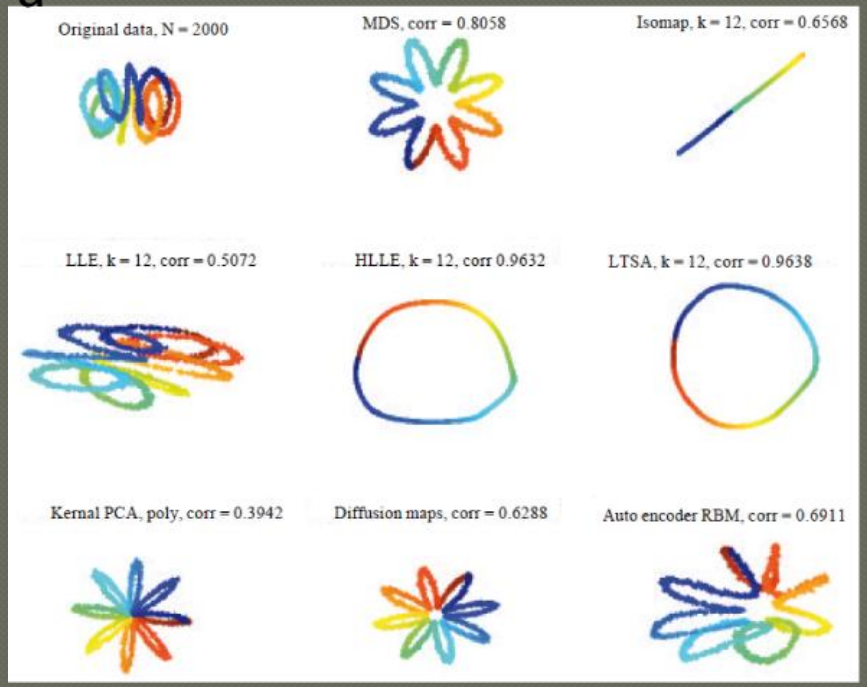
Independence
(i.e.: Independent Component Analysis)

Decoster, 1998

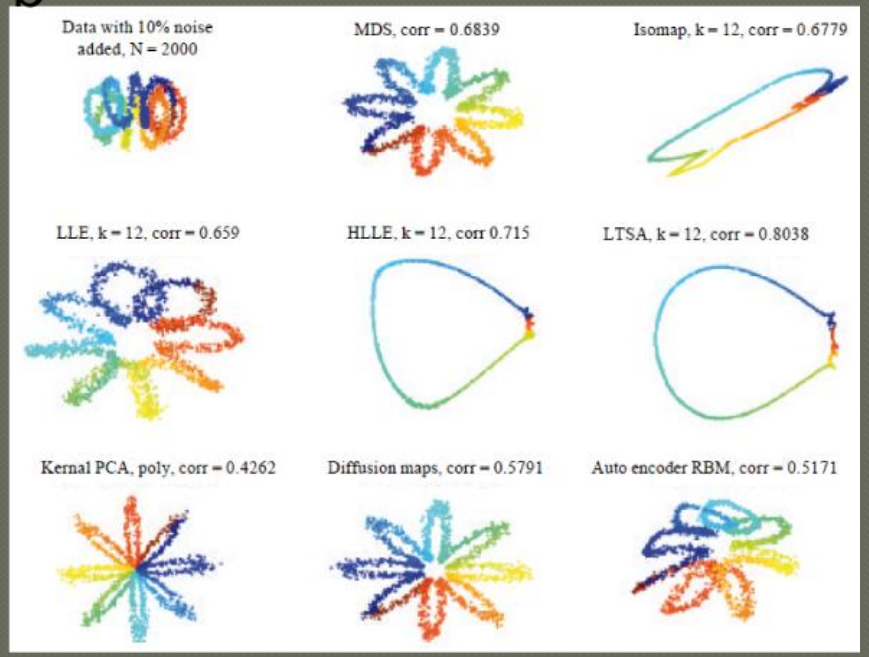


Effect of noise

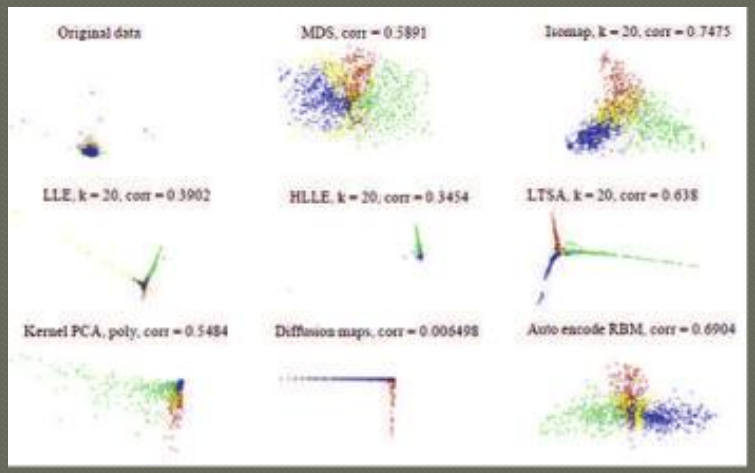
a

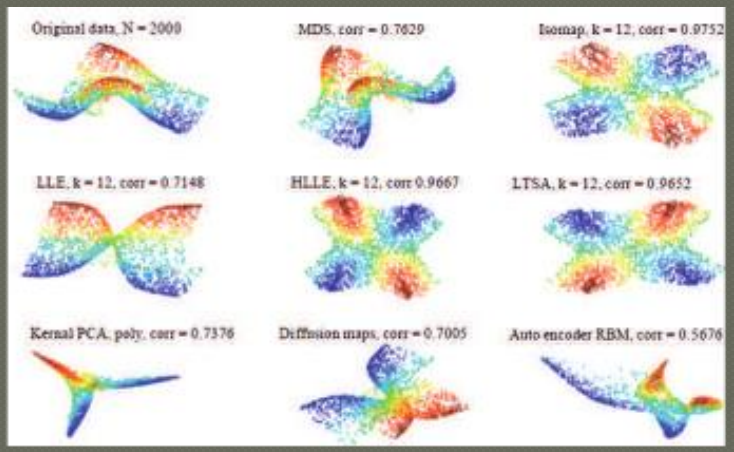


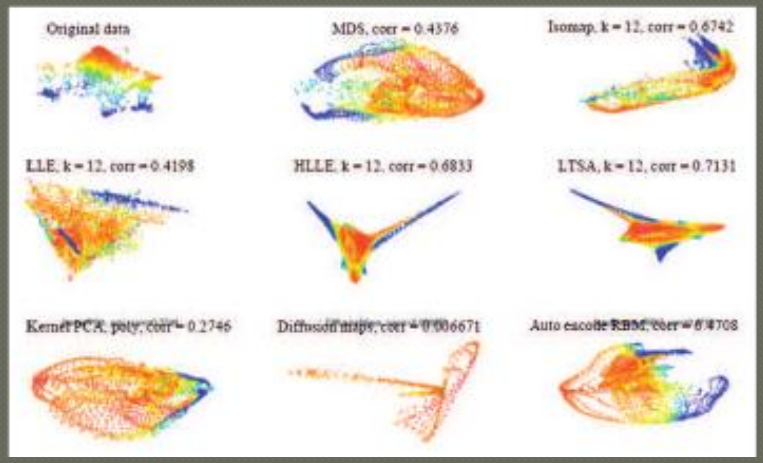
b

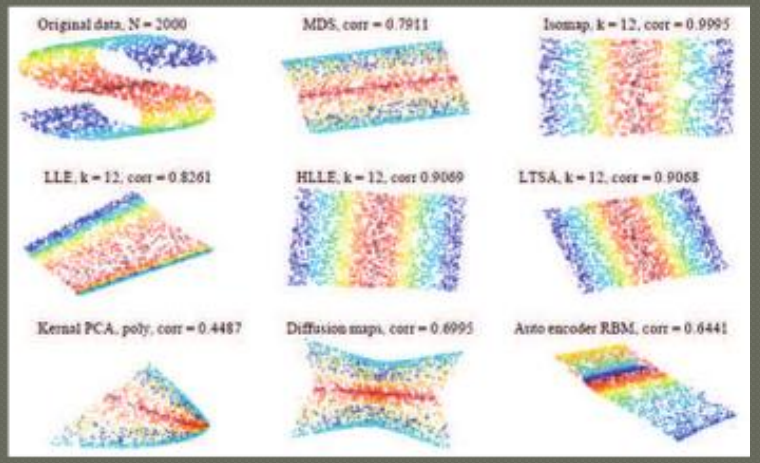


A great diversity of results given by each dimensionality reduction technique

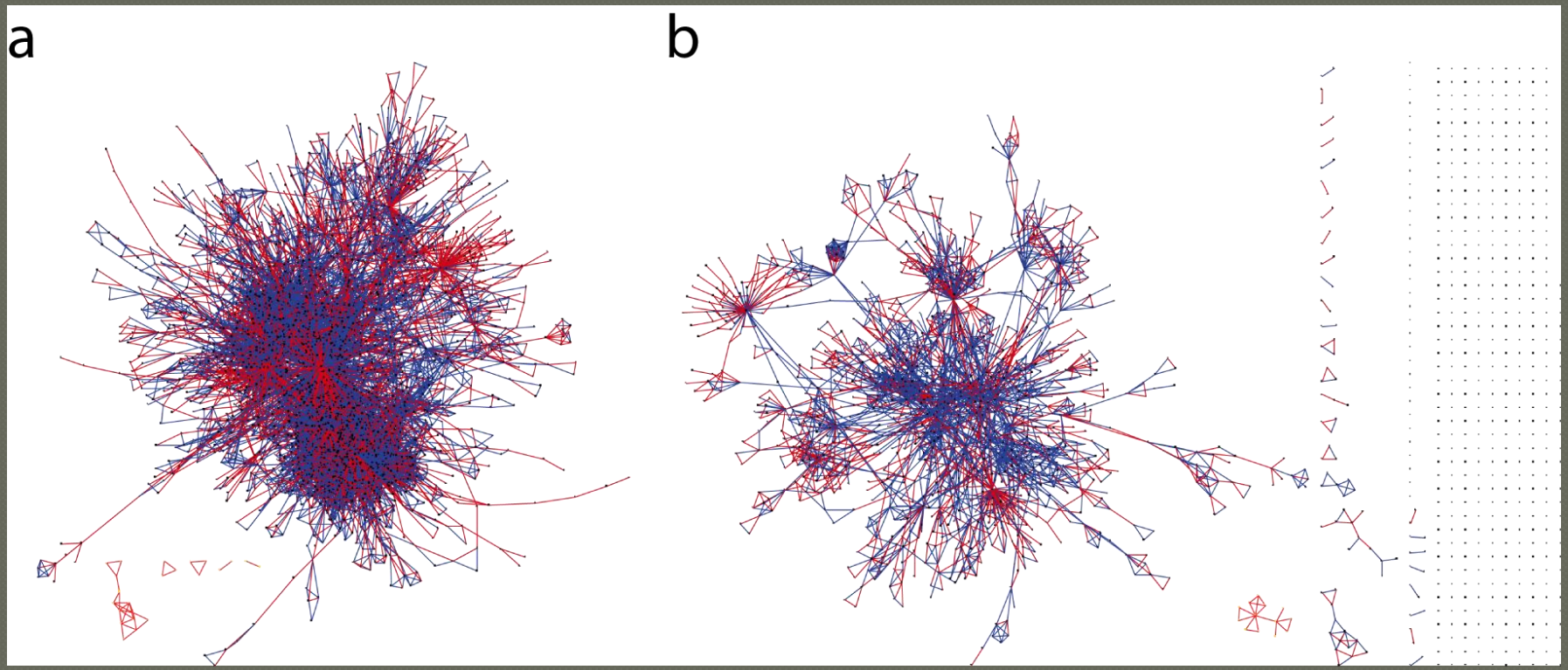






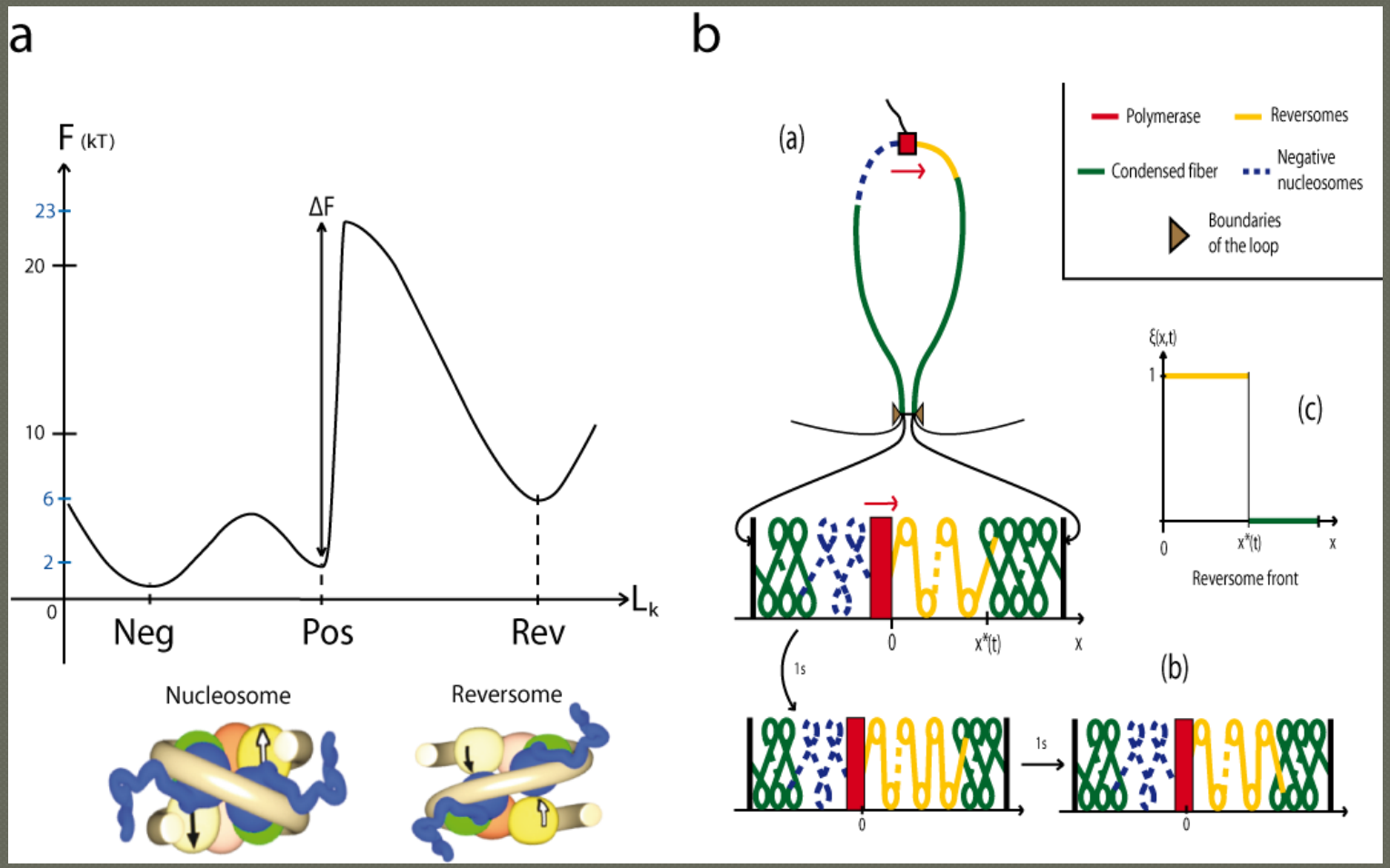


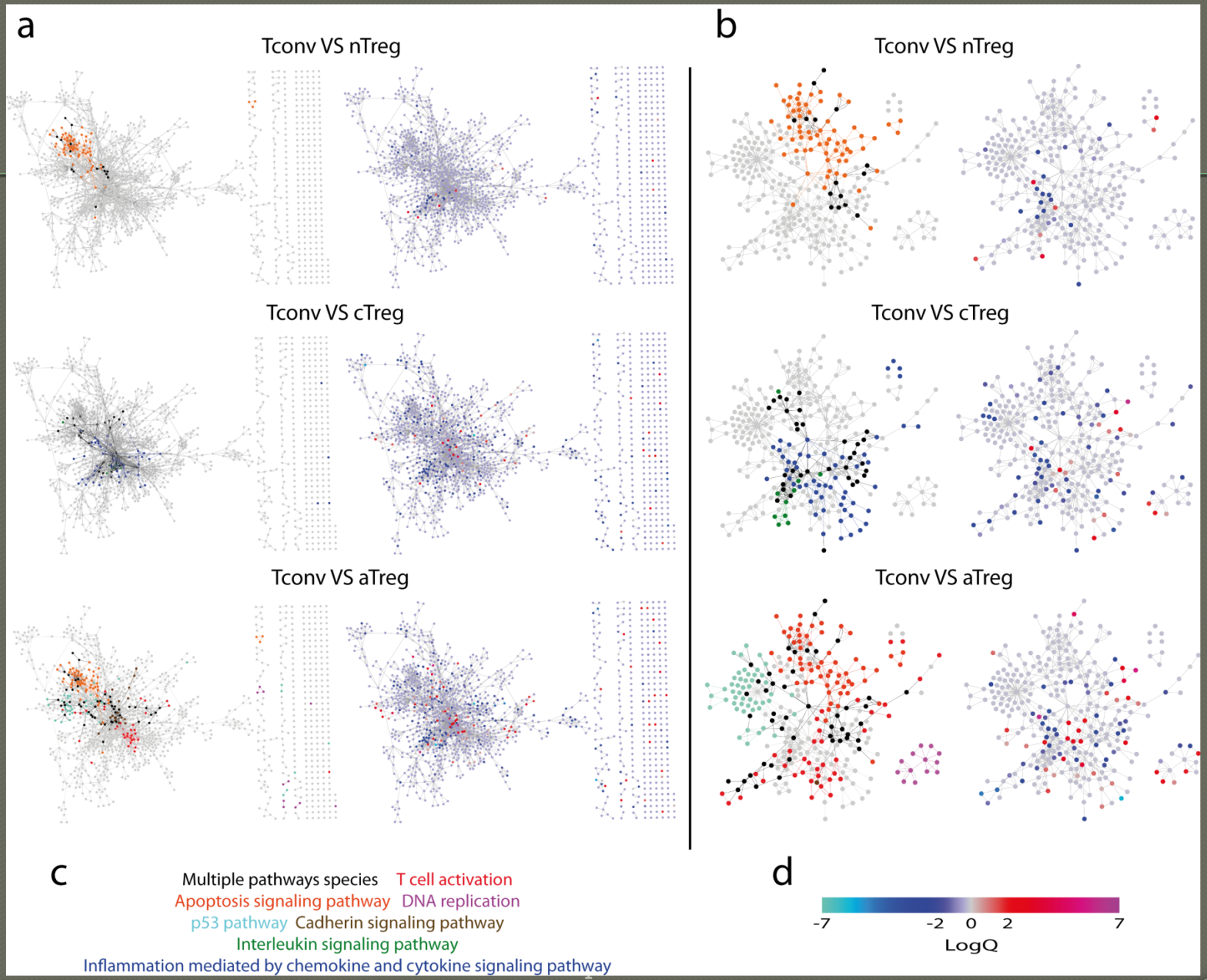
Tsai F., Journal of Artificial Intelligence, 2010

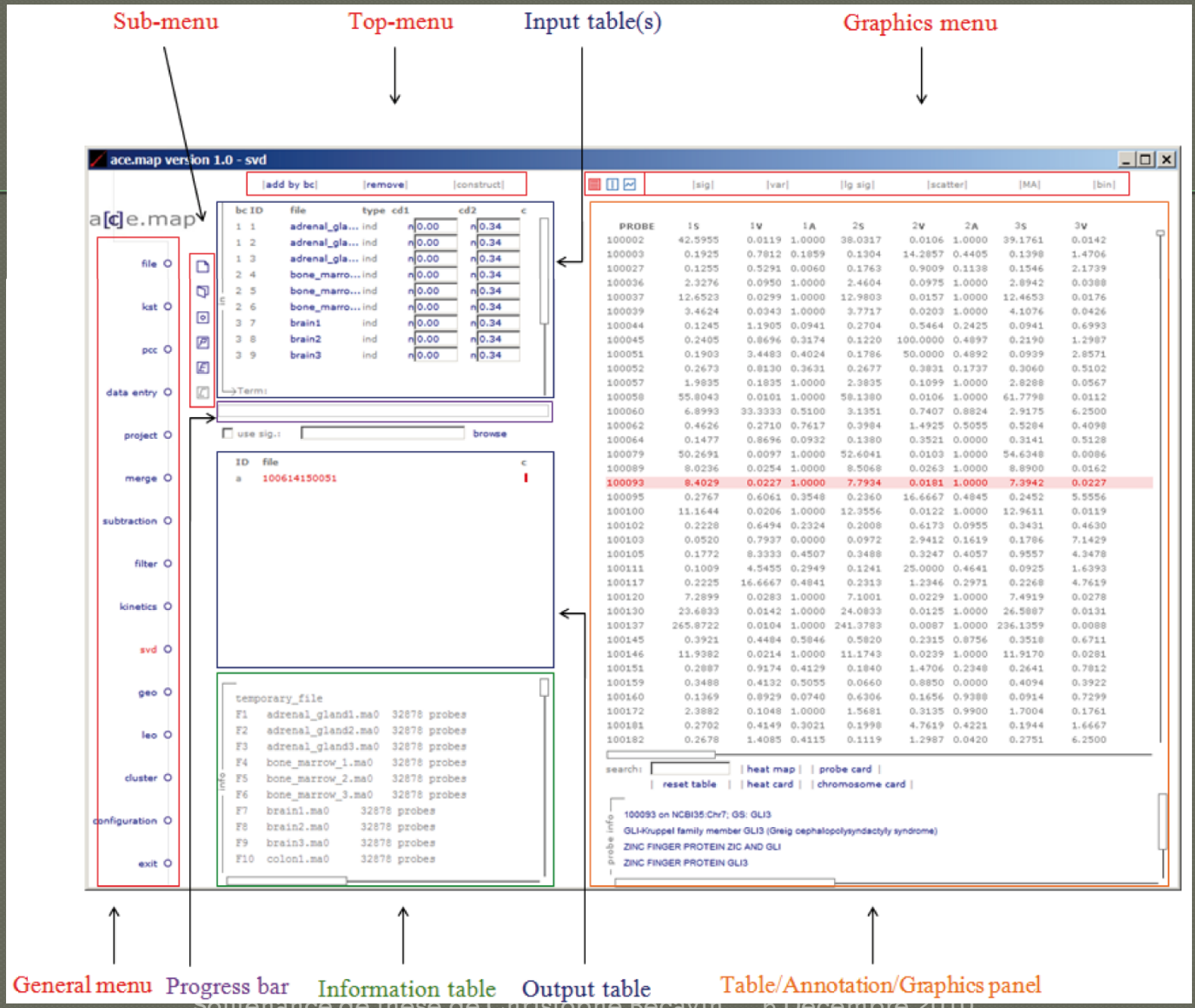

a
b

Chromatin during transcription

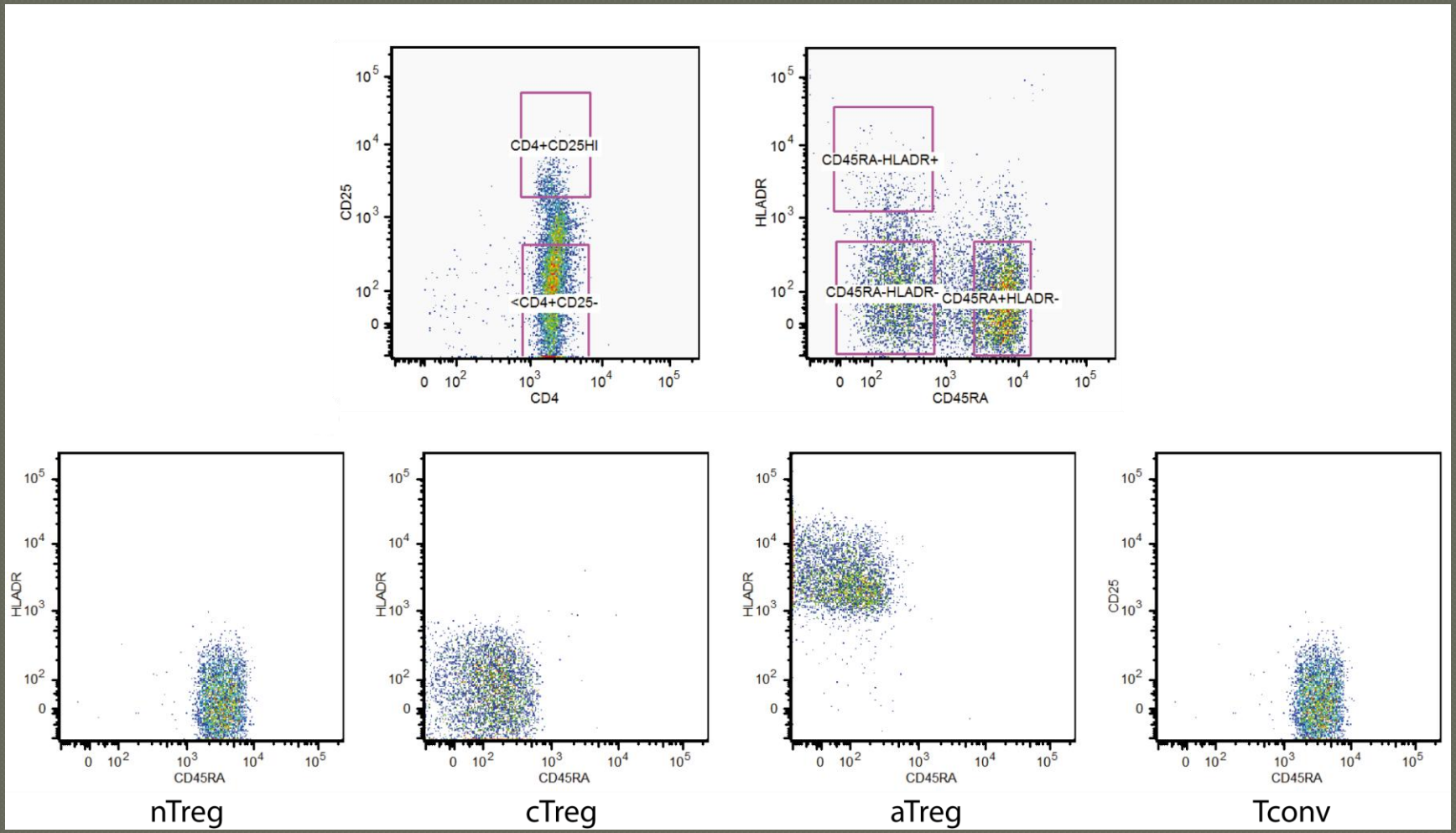
Bécavin et al., Biophysical Journal, 2009

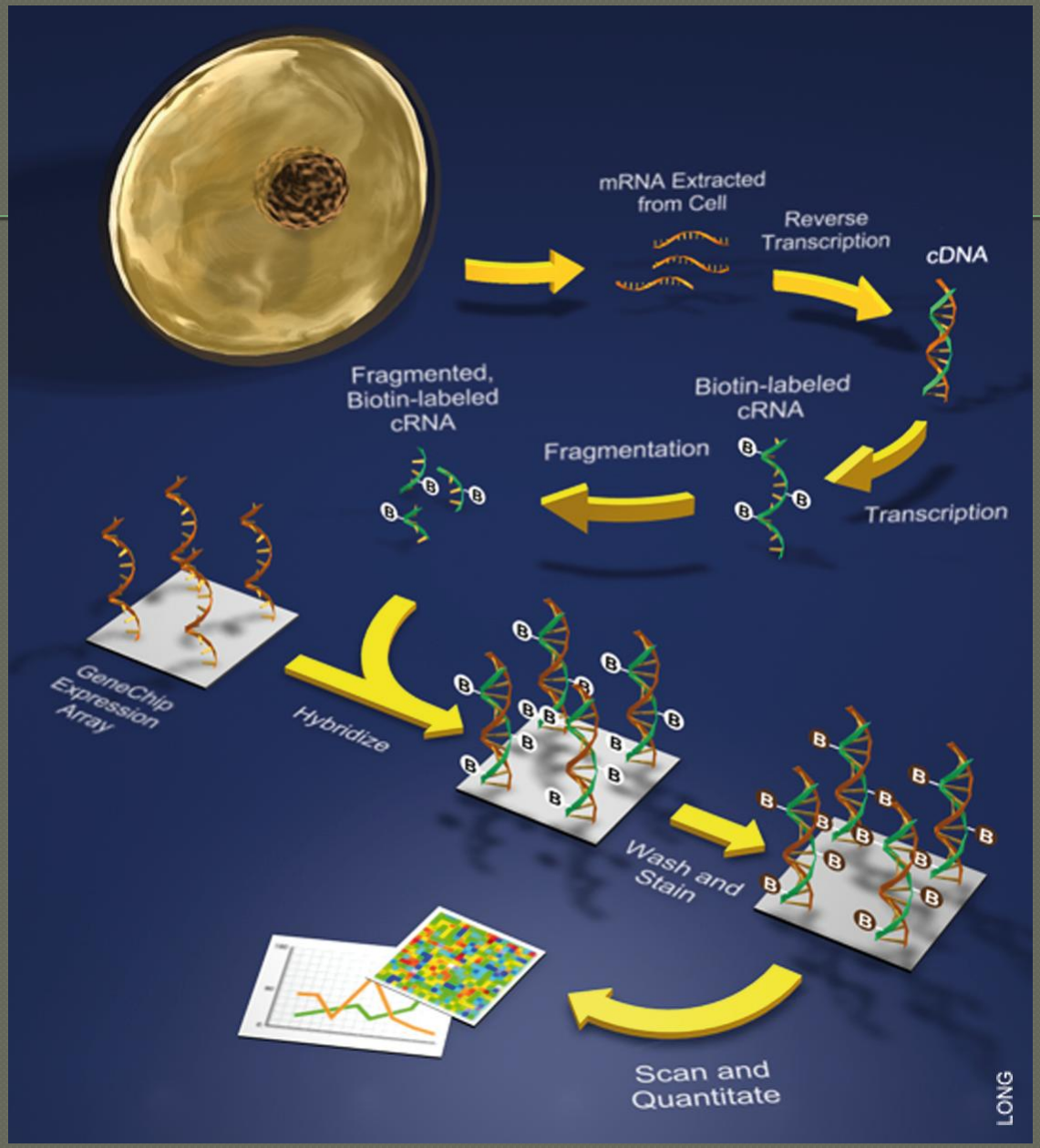



a
Tconv VS nTreg
Tconv VS cTreg
Tconv VS aTreg
b
Tconv VS nTreg
Tconv VS cTreg
Tconv VS aTreg
c
Multiple pathways species
T cell activation
Apoptosis signaling pathway
DNA replication
p53 pathway
Cadherin signaling pathway
Interleukin signaling pathway
Inflammation mediated by chemokine and cytokine signaling pathway
d
-7
-2
0
2
7
LogQ


Sub-menu
Top-menu
Input table(s)
Graphics menu
ace.map version 1.0 - svd
[add by bc]
[remove]
[construct]
[sig]
[var]
[lg sig]
[scatter]
[MA]
[bin]
a[c]e.map
file
kat
pcc
data entry
project
merge
subtraction
filter
kinetics
svd
geo
leo
cluster
configuration
exit
bc ID file type cd1 cd2 c
1 1 adrenal_gla... ind n 0.00 n 0.34
1 2 adrenal_gla... ind n 0.00 n 0.34
1 3 adrenal_gla... ind n 0.00 n 0.34
2 4 bone_marro... ind n 0.00 n 0.34
2 5 bone_marro... ind n 0.00 n 0.34
2 6 bone_marro... ind n 0.00 n 0.34
3 7 brain1 ind n 0.00 n 0.34
3 8 brain2 ind n 0.00 n 0.34
3 9 brain3 ind n 0.00 n 0.34
Term:
use sig.:
browse
ID file c
a 100614150051
temporary_file
F1 adrenal_gland1.ma0 32878 probes
F2 adrenal_gland2.ma0 32878 probes
F3 adrenal_gland3.ma0 32878 probes
F4 bone_marrow_1.ma0 32878 probes
F5 bone_marrow_2.ma0 32878 probes
F6 bone_marrow_3.ma0 32878 probes
F7 brain1.ma0 32878 probes
F8 brain2.ma0 32878 probes
F9 brain3.ma0 32878 probes
F10 colon1.ma0 32878 probes
info
PROBE 1S 1V 1A 2S 2V 2A 3S 3V
100002 42.5955 0.0119 1.0000 38.0317 0.0106 1.0000 39.1761 0.0142
100003 0.1925 0.7812 0.1859 0.1304 14.2857 0.4405 0.1398 1.4706
100027 0.1255 0.5291 0.0060 0.1763 0.9009 0.1138 0.1546 2.1739
100036 2.3276 0.0950 1.0000 2.4604 0.0975 1.0000 2.8942 0.0388
100037 12.6523 0.0299 1.0000 12.9803 0.0157 1.0000 12.4653 0.0176
100039 3.4624 0.0343 1.0000 3.7717 0.0203 1.0000 4.1076 0.0426
100044 0.1245 1.1905 0.0941 0.2704 0.5464 0.2425 0.0941 0.6993
100045 0.2405 0.8696 0.3174 0.1220 100.0000 0.4897 0.2190 1.2987
100051 0.1903 3.4483 0.4024 0.1786 50.0000 0.4892 0.0939 2.8571
100052 0.2673 0.8130 0.3631 0.2677 0.3831 0.1737 0.3060 0.5102
100057 1.9835 0.1835 1.0000 2.3835 0.1099 1.0000 2.8288 0.0567
100058 55.8043 0.0101 1.0000 58.1380 0.0106 1.0000 61.7798 0.0112
100060 6.8993 33.3333 0.5100 3.1351 0.7407 0.8824 2.9175 6.2500
100062 0.4626 0.2710 0.7617 0.3984 1.4925 0.5055 0.5284 0.4098
100064 0.1477 0.8696 0.0932 0.1380 0.3521 0.0000 0.3141 0.5128
100079 50.2691 0.0097 1.0000 52.6041 0.0103 1.0000 54.6348 0.0086
100089 8.0236 0.0254 1.0000 8.5068 0.0263 1.0000 8.8900 0.0162
100093 8.4029 0.0227 1.0000 7.7934 0.0181 1.0000 7.3942 0.0227
100095 0.2767 0.6061 0.3548 0.2360 16.6667 0.4845 0.2452 5.5556
100100 11.1644 0.0206 1.0000 12.3556 0.0122 1.0000 12.9611 0.0119
100102 0.2228 0.6494 0.2324 0.2008 0.6173 0.0955 0.3431 0.4630
100103 0.0520 0.7937 0.0000 0.0972 2.9412 0.1619 0.1786 7.1429
100105 0.1772 8.3333 0.4507 0.3488 0.3247 0.4057 0.9557 4.3478
100111 0.1009 4.5455 0.2949 0.1241 25.0000 0.4641 0.0925 1.6393
100117 0.2225 16.6667 0.4841 0.2313 1.2346 0.2971 0.2268 4.7619
100120 7.2899 0.0283 1.0000 7.1001 0.0229 1.0000 7.4919 0.0278
100130 23.6833 0.0142 1.0000 24.0833 0.0125 1.0000 26.5887 0.0131
100137 265.8722 0.0104 1.0000 241.3783 0.0087 1.0000 236.1359 0.0088
100145 0.3921 0.4484 0.5846 0.5820 0.2315 0.8756 0.3518 0.6711
100146 11.9382 0.0214 1.0000 11.1743 0.0239 1.0000 11.9170 0.0281
100151 0.2887 0.9174 0.4129 0.1840 1.4706 0.2340 0.2641 0.7812
100159 0.3488 0.4132 0.5055 0.0660 0.8850 0.0000 0.4094 0.3922
100160 0.1369 0.8929 0.0740 0.6306 0.1656 0.9388 0.0914 0.7299
100172 2.3882 0.1048 1.0000 1.5681 0.3135 0.9900 1.7004 0.1761
100181 0.2702 0.4149 0.3021 0.1998 4.7619 0.4221 0.1944 1.6667
100182 0.2678 1.4085 0.4115 0.1119 1.2987 0.0420 0.2751 6.2500
search:
heat map
probe card
reset table
heat card
chromosome card
probe info
100093 on NCBI35:Chr7; GS: GLI3
GLI-Kruppel family member GLI3 (Greig cephalopolysyndactyly syndrome)
ZINC FINGER PROTEIN ZIC AND GLI
ZINC FINGER PROTEIN GLI3
General menu
Progress bar
Information table
Output table
Table/Annotation/Graphics panel

Treg purification using FACS




mRNA Extracted from Cell
Reverse Transcription
cDNA
Biotin-labeled cRNA
Transcription
Fragmentation
Fragmented, Biotin-labeled cRNA
GeneChip Expression Array
Hybridize
Wash and Stain
Scan and Quantitate
LONG